THE ANNALS *of* MATHEMATICAL STATISTICS

FOUNDED AND EDITED BY H. C. CARVER, 1930-1938

EDITED BY S. S. WILKS, 1938-1949

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. 22, No. 2 — June, 1951

THE ANNALS OF MATHEMATICAL STATISTICS

Published quarterly by the Institute of Mathematical Statistics in March, June, September and December at Baltimore, Maryland

INSTITUTE OF MATHEMATICAL STATISTICS

General Office: Business Administration Building, University of Michigan, Ann Arbor, Michigan
C. H. Fischer, Secretary-Treasurer
This address should be used for all communications concerning membership, subscriptions, changes of address, back numbers, etc., but *not* for editorial correspondence. Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue.

Editorial Office: Department of Mathematical Statistics, Columbia University, New York 27, New York
T. W. Anderson, Editor
Manuscripts should be submitted to this address; each manuscript should be typewritten, double-spaced with wide margins, and the original copy should be submitted (preferably with one additional copy). Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be professionally drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae. Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

Subscription Price: 1951: $10 per year in Western Hemisphere; $5 elsewhere; single issues $3; back numbers $10 per volume, $3 per issue.
1952: $12 per year in U. S. and Canada; $10 elsewhere.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC., BALTIMORE, MARYLAND, U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the act of March 3, 1879

CONSISTENCY AND UNBIASEDNESS OF CERTAIN NONPARAMETRIC TESTS

BY E. L. LEHMANN

University of California, Berkeley

1. Summary. It is shown that there exist strictly unbiased and consistent tests for the univariate and multivariate two- and k-sample problem, for the hypothesis of independence, and for the hypothesis of symmetry with respect to a given point. Certain new tests for the univariate two-sample problem are discussed. The large sample power of these tests and of the Mann-Whitney test are obtained by means of a theorem of Hoeffding. There is a discussion of the problem of tied observations.

2. Introduction. The purpose of the present paper is to investigate the existence and various properties of strictly unbiased and of consistent tests for testing certain nonparametric hypotheses. The problems that will be considered are the two-sample and k-sample problem, the hypothesis of independence and the hypothesis of symmetry with respect to a given point.

A sequence of tests is said to be consistent against a certain class of alternatives if for each alternative the power of the test tends to one as the sample sizes tend to infinity. A test will be said to be strictly unbiased if the power for each alternative exceeds the level of significance.

Consistency being a rather weak property, which one would expect most sequences of tests to satisfy for the class of alternatives for which they are designed, it is important to obtain some more detailed information concerning the power of the various tests under consideration. Because of the tremendous variety of the alternatives it seems fairly hopeless to get a comprehensive view of the achievements of most tests when the samples are small. This in spite of the fact that it is occasionally possible to write down the power explicitly (for example in the simplest cases of the tests discussed by Mathisen [1]). On the other hand, the large sample distribution of a number of test statistics may be found by means of the asymptotic theorems of Hoeffding [2]. Asymptotically, the power then usually involves only a few parameters and a large sample comparison of various different tests becomes possible.

3. Two-sample problem: specific classes of alternatives. We shall discuss in detail only one of the problems mentioned, the two-sample problem, and indicate only briefly certain extensions to the other problems. In the two-sample problem one is given independent samples $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ from populations with unknown cumulative distribution functions F and G respectively, and it is desired to test the hypothesis $F = G$. In this connection various classes of alternatives are possible.

It may, for example, be known that unless $F = G$, the Y's tend to be larger than the X's. For this problem it has been proposed as a test to consider the

number of pairs X_i, Y_j for which $X_i < Y_j$, and to reject the hypothesis if this number is too large. This test was proved consistent by Mann and Whitney [3] against the alternatives that

$$F(t) > G(t) \quad \text{for all} \quad t. \tag{3.1}$$

Actually their proof shows that the test is consistent[1] against all alternatives for which $P(Y_j > X_i) > \frac{1}{2}$.

We shall now prove that this test is also unbiased against the alternatives satisfying (3.1).[2] This is true not only for this test but also for those proposed by Thompson [4] and for tests based on randomisation of such statistics as $\bar{y} - \bar{x}$. In fact we have

THEOREM 3.1. *Let w be any similar region for testing $H: F = G$ on the basis of $X_1, \cdots, X_m; Y_1, \cdots, Y_n$. Suppose w is such that $(x_1, \cdots, x_m; z_1, \cdots, z_n) \varepsilon w$ and $y_i \geqq z_i$ for $i = 1, \cdots, n$ implies $(x_1, \cdots, x_m; y_1, \cdots, y_n) \varepsilon w$. Then the test is unbiased against all continuous alternatives F, G satisfying* (3.1).

PROOF. Suppose that $X_1, \cdots, X_m, Y_1, \cdots, Y_n$ are independent and all have the same c.d.f. F and that G is such that (3.1) holds. Then we shall construct $Y_i = f(Z_i)$ such that $Y_i > Z_i$ for $i = 1, \cdots, n$ and such that the Y's have c.d.f. G. Thus the probability of $(X_1, \cdots, X_m; Z_1, \cdots, Z_n) \varepsilon w$ equals the level of significance, α say, while the probability of $(X_1, \cdots, X_m; Y_1, \cdots, Y_n) \varepsilon w$ equals the power of the test against the alternative (F, G). But since $Y_i > Z_i$ for all i, the test rejects for the X's and Y's whenever it rejects for the X's and Z's, and hence the power is $\geqq \alpha$.

The function f is easily defined by the equation

$$G(f(z)) = F(z).$$

(When this does not define $f(z)$ uniquely, it does not matter which of the possible definitions is used.) That $y = f(z) > z$ follows from assumption (3.1).

The theorem as stated refers only to tests in which no randomisation is allowed, but the extension to randomised tests is immediate. Also, as we shall show later, the assumption of continuity of F and G may be omitted.

Theorem 3.1 may be used also to widen the applicability of the tests to which it refers. So far, we have taken the hypothesis to state that X and Y have the same distribution. This formulation may arise, for example, when one is faced with the question whether a treatment, known to be harmless, has a beneficial effect: Either it has no effect so that $F = G$, or it has a good effect. If, on the other hand, the comparison is between two different treatments one may wish to test hypothesis H' that Y tends to be smaller than X, against the alternatives that it tends to be larger. The hypothesis would then be

$$H': F(t) \leqq G(t) \quad \text{for all} \quad t.$$

[1] This was also noticed by van Dantzig who points it out in a paper "On the consistency and the power function of Wilcoxon's two sample test," to be published in *Proc. Roy. Inst. Acad. Sci.*, 1951.

[2] For alternatives (F, G) differing only in location this was proved by Van der Vaart [26].

There is of course no nontrivial similar region for this problem, however any region w satisfying the condition of Theorem 3.1 and such that $P(w) = \alpha$ whenever $F \neq G$ clearly will be of size α for testing H' i.e. $P(w)$ will be $\leqq \alpha$ whenever $F(t) \leqq G(t)$ for all t.

Returning to the Mann-Whitney test, let us define V by

$$mnV = \text{number of pairs } X_i, Y_j \text{ with } X_i < Y_j. \tag{3.2}$$

It was shown by Mann and Whitney that V is asymptotically normally distributed when $F = G$ and $m, n \to \infty$ in an arbitrary manner. From a result of Hoeffding (Theorem 7.3 of [2]) it follows that asymptotic normality holds also when $F \neq G$ provided m/n remains constant as $m, n \to \infty$.

We shall apply Hoeffding's theorem to prove asymptotic normality not only of V, but of a large class of statistics connected with the two-sample problem. We begin by stating Hoeffding's theorem, somewhat specialised and with slight changes of notation:

Let $Z_1, \cdots, Z_n$ be independently, identically distributed chance vectors with real components, let $s \leqq n$ and let $\phi(Z_1, \cdots, Z_s)$ be a real valued symmetric function of its s arguments such that $E[\phi(Z_1, \cdots, Z_s)]^2 < \infty$, and let us write

$$E\phi(Z_1, \cdots, Z_s) = \theta.$$

Let

$$U_n = \binom{n}{s}^{-1} \Sigma\phi(Z_{\alpha_1}, \cdots, Z_{\alpha_s}),$$

where the summation extends over all subscripts $1 \leqq \alpha_1 < \cdots < \alpha_s \leqq n$, and let

$$U_n' = U_n + R_n,$$

where R_n is a random variable for which

$$E(nR_n^2) \to 0 \quad \text{as} \quad n \to \infty.$$

Then $\sqrt{n}(U_n' - \theta)$ is asymptotically normally distributed. Further, if we put

$$\psi(z_1) = E[\phi(z_1, Z_2, \cdots, Z_s) - \theta],$$

the limiting distribution of $\sqrt{n}(U_n' - \theta)$ is nondegenerate provided $E[\psi(Z_1)]^2 > 0$.

We can now state

THEOREM 3.2. *Let $X_1, \cdots, X_m$; $Y_1, \cdots, Y_n$ be independently distributed with* c.d.f.'s *F, G respectively. Let $t(x_1, \cdots, x_r, y_1, \cdots, y_r)$ be symmetric in the x's alone and in the y's alone. Suppose that*

$$Et(X_1, \cdots, X_r, Y_1, \cdots, Y_r) = \theta(F, G) = \theta,$$

$$E[t(X_1, \cdots, X_r, Y_1, \cdots, Y_r)]^2 = M < \infty.$$

Let $m/n = c$, and let n be sufficiently large so that $r \leqq m, n$. Define

$$U_n' = \binom{m}{r}^{-1}\binom{n}{r}^{-1} \Sigma t(X_{\alpha_1}, \cdots, X_{\alpha_r}, Y_{\beta_1}, \cdots, Y_{\beta_r}),$$

where the summation is extended over all subscripts $1 \leqq \alpha_1 < \cdots < \alpha_r \leqq m$; $1 \leqq \beta_1 < \cdots < \beta_r \leqq n$. *Then, as* $n \to \infty$, $\sqrt{n}(U'_n - \theta)$ *is asymptotically normally distributed.*

PROOF. For the sake of simplicity we shall give the proof only in the case $m = n$. Let $Z_i = (X_i, Y_i)$ and define

$$\phi(Z_1, \cdots, Z_{2r}) = \binom{2r}{r}^{-1} \Sigma t(X_{i_1}, \cdots, X_{i_r}, Y_{j_1}, \cdots, Y_{j_r})$$

summed over all sets of indices for which $(i_1 < \cdots < i_r, j_1 < \cdots < j_r)$ is a permutation of $(1, \cdots, 2r)$. Further let

$$U_n = \binom{n}{2r}^{-1} \Sigma\phi(Z_{\gamma_1}, \cdots, Z_{\gamma_{2r}})$$

summed over all γ's such that $1 \leqq \gamma_1 < \cdots < \gamma_{2r} \leqq n$.

Clearly $\binom{n}{r}^2 U'_n$ is the sum of all possible t-terms, while $\binom{2r}{r}\binom{n}{2r}U_n$ is the sum of only those t-terms in which the X's and Y's have no common subscript. Hence, since $\binom{n}{2r}\binom{2r}{r} = \binom{n}{r}\binom{n-r}{r}$, we have

$$U'_n = \binom{n-r}{r}\binom{n}{r}^{-1} U_n + \binom{n}{r}^{-2} W_n,$$

where W_n is a sum of $\left[\binom{n}{r}\binom{n}{r} - \binom{n}{r}\binom{n-r}{r}\right]$ t-terms, and we can write

$$U'_n = U_n + D_n,$$

where

$$D_n = \left[\binom{n-r}{r} - \binom{n}{r}\right]\binom{n}{r}^{-1} U_n + \binom{n}{r}^{-2} W_n.$$

Since for any real numbers $t_1, \cdots, t_k$ we have $(t_1 + \cdots + t_k)^2 \leqq k(t_1^2 + \cdots + t_k^2)$ we see that

$$E(D_n^2) \leqq \left[\binom{n}{r} - \binom{n-r}{r}\right]^2 \binom{n}{r}^{-2} M + \left[\binom{n}{r} - \binom{n-r}{r}\right]\binom{n}{r}^{-3} M.$$

But, as $n \to \infty$, $\sqrt{n}\left[\binom{n}{r} - \binom{n-r}{r}\right]\binom{n}{r}^{-1} \to 0$. Hence $E(nD_n^2) \to 0$ and the result follows.

Let us now consider the application of this theorem to the Mann-Whitney statistic. We define

$$t(x, y) = \begin{cases} 1 & \text{if} \quad y > x, \\ 0 & \text{if} \quad y \leqq x. \end{cases}$$

Then $U'_n = V_{m,n}$ and asymptotic normality follows since $E\varphi^2(X, Y) \leqq 1$. It remains to check under what conditions $E\psi^2(Z_1) > 0$. Since we have $s = 2r = 2$,

$$2\psi(z_1) = P(Y_2 > x_1) + P(y_1 > X_2) - 2P(Y > X).$$

Hence $E\psi^2(Z) = 0$ is equivalent to $F(Y) - G(X) =$ constant with probability 1, or $P(Y > x) + P(y > X) =$ constant except on a set of points (x, y) that has probability zero. It is easy to see that this is satisfied if and only if $P(Y > X)$ is 1 or 0.

So far we have considered the hypothesis $H: F = G$ against the alternatives that the Y's tend to be larger than the X's. As a second example we shall consider testing H, or even the wider hypothesis H' that F and G differ only in location (i.e., that $F(x) = G(x + d)$ for some d), against the alternative that the Y's are more spread out than the X's (in a sense to be defined below). In analogy with the Mann-Whitney test let $W_{m,n}$ be the proportion of quadruples X_i, X_j, Y_k, Y_l for which $| Y_l - Y_k | > | X_j - X_i |$. We reject H if $W_{m,n}$ is too large. This test is unbiased against all alternatives (F, G) for which $F(x_1) = G(y_1)$, $F(x_2) = G(y_2)$ implies $| x_1 - x_2 | < | y_1 - y_2 |$. The test is consistent against the wider class of alternatives for which $P(| Y' - Y | > | X' - X |) > \frac{1}{2}$ where X, X', Y, Y' are independently distributed with c.d.f. F, G, respectively. The proof of unbiasedness is quite analogous to the one given previously, and we shall therefore omit it.

We shall however indicate the proof of consistency, and refer in this connection to the closely related remarks by Hoeffding [5] on the construction of consistent sequences of tests.

We first state for reference the following trivial

Lemma 3.1. *Let $\theta = f(F, G)$ be a real valued function such that $f(F, F) = \theta_0$ for all (F, F) in a class $\mathfrak{C}_0$. Let $T_{m,n} = t_{m,n}(X_1, \cdots, X_m, Y_1, \cdots, Y_n)$ be a sequence of real valued statistics such that $T_{m,n}$ tends to θ in probability as* min $(m, n) \to \infty$. *Suppose that $f(F, G) > \theta_0 (\neq \theta_0)$ for all (F, G) in a class $\mathfrak{C}_1$. Then the sequence of tests which reject when $T_{m,n} - \theta_0 > C_{m,n}$ (when $| T_{m,n} - \theta_0 | > C'_{m,n}$) is consistent for testing $H: \mathfrak{C}_0$ at every fixed level of significance against the alternatives $\mathfrak{C}_1$.*

For proof one need only to notice that a fixed level of significance $\neq 0$ implies that $C_{m,n} \to 0$ $(C'_{m,n} \to 0)$ as $m, n \to \infty$.

In the applications we have in mind, $E(T_{m,n})$ is usually independent of m and n, and is easy to find. On the other hand some work is required to determine $\sigma^2(T_{m,n})$. It is therefore of interest to notice that the evaluation of $\sigma^2(T_{m,n})$ is frequently not necessary to prove consistency. To this end we shall state the following lemma, which is a generalisation of a theorem of Halmos [6], and which follows easily from Theorem 5.1 of [7]. A simple proof will be given in [8].

Lemma 3.2 (Lehmann-Scheffé). *Let $f(F, G)$ be a real valued function defined for all continuous* c.d.f.'s *F and G. There exists at most one function $t_{m,n}$ such that $t_{m,n}(X_1, \cdots, X_m, Y_1, \cdots, Y_n)$ is symmetric in the first m and in the last n arguments and is an unbiased estimate of $f(F, G)$ for all continuous (or even ab-*

solutely continuous) c.d.f.'s F, G. *If such a function* $t_{m,n}$ *exists, (and has finite variance), it has among all unbiased estimates of* $f(F, G)$ *uniformly smallest variance.*

For the application to be made here we need the slightly stronger statement that the conclusion of the Lemma remains valid if $t_{m,n}(X_1, \cdots, X_m; Y_1, \cdots, Y_n)$ is an unbiased estimate of $f(F, G)$ for all continuous c.d.f.'s F, G for which

$$P(|Y' - Y| > |X' - X|) > \tfrac{1}{2}.$$

This generalization follows immediately from the proof of the Lemma given in [8].

The proof of consistency of the proposed test is now immediate. For let $W'_{m,n}$ be the proportion of quadruples for which the Y's are further apart than the X's among the independent quadruples $X_1, X_2, Y_1, Y_2; X_3, X_4, Y_3, Y_4; \cdots$. Then

$$E(W'_{m,n}) = E(W_{m,n}) = P(|Y' - Y| > |X' - X|),$$

and hence by Lemma 3.2,

$$\sigma^2(W_{m,n}) \leqq \sigma^2(W'_{m,n}). \tag{3.3}$$

But $\sigma^2(W'_{m,n})$ obviously tends to zero as $m, n \to \infty$.

We remark finally that the large sample distribution of $W_{m,n}$, by Theorem 3.1, is again normal. Degeneracy occurs only if either F or G are one-point distributions.

As a last problem in this section we shall consider the hypothesis $F = G$ against the combined class of alternatives that the Y's are larger than the X's or more spread out. In such a problem it seems important not only to decide whether F and G are equal but, in case the hypothesis is rejected, for which of the two possible reasons it is rejected or whether it is rejected for both of them. (See in this connection the discussion by Berkson [9]). Thus one is really dealing with a multidecision problem. One must decide between

d_0: Accepting the hypothesis $H: F = G$,
d_1: Rejecting H for the reason that the Y's are larger than the X's,
d_2: Rejecting H for the reason that the Y's are more spread out than the X's,
d_3: Rejecting H for both reasons.

It is desired to find a decision procedure under which the probability of taking decision d_0 is $1 - \alpha$ when $F = G$ while the probability of taking the appropriate of the decisions d_1, d_2, d_3 when the hypothesis is false tends to 1 as the sample sizes tend to infinity. Let us recall the statistics $V_{m,n}$ and $W_{m,n}$ introduced in connection with the previous problems and let us denote $E(V_{m,n})$ and $E(W_{m,n})$ by θ and η respectively. One may then accept H when $V_{m,n} \leqq a_{m,n}$, $W_{m,n} \leqq b_{m,n}$, or take one of the remaining three decisions according as to which one of the three complementary inequalities holds. The constants $a_{m,n}$ and $b_{m,n}$ are not completely determined by the equation

$$P(V_{m,n} \leqq a_{m,n}, W_{m,n} \leqq b_{m,n} \mid F = G) = \alpha.$$

One may specify some additional restriction, such as

$$P(V_{m,n} \leqq a_{m,n} \mid F = G) = P(W_{m,n} \leqq b_{m,n} \mid F = G).$$

It is easy to prove that the above procedure has the consistency property asked for. This follows from Lemmas 3.1 and 3.2 generalised to the case that the function $f(F, G)$ of these lemmas is vector valued instead of real valued. The function $t_{m,n}$ of Lemma 3.2 is then also vector valued and instead of its variance one may consider its ellipsoid of concentration (see [10] and Theorem 5.2 of [7]). In the present case we notice that $(V_{m,n}, W_{m,n})$ is a symmetric estimate of (θ, η) and hence has a uniformly smallest ellipsoid of concentration. But one can easily construct unbiased estimates of θ and η based on independent samples, whose concentration ellipsoid has both axes tending to zero as the sample sizes increase indefinitely and so consistency can be proved by the device used after Lemma 3.2.

4. Two-sample problem: general class of alternatives. In the present section we shall consider the problem of testing the hypothesis $F = G$ against the class of all continuous alternatives $F \neq G$. One might argue that this should not be treated as a hypothesis-testing problem. For Berkson's argument seems to apply: The question is not only whether or not the hypothesis is true. If it is false, it is necessary to decide what alternative hypothesis is correct. While in some situations, this criticism seems to be valid, there are others in which it does not seem to apply.

The two-sample problem may arise in the following two quite different settings.

A: Two production processes, treatments or populations are available, and it is desired to decide whether one is better than the other. In this case the populations F and G are in competition, and the main problem is that of ranking them. Here the notion of such a ranking automatically suggests some specific class or classes of alternatives to the hypothesis that the populations do not differ.

B: The two populations coexist. There is no question of which is preferable, but we wish to know whether the two can be treated as one. One may, for example, want to know whether the output of two different machines can be treated as a uniform product, or whether data obtained under two different experimental setups or by two different investigators may be pooled. These problems really are two-decision problems: The data can or can not be pooled. An explanation of why they can not be pooled is not necessarily of interest.

In connection with the present problem Wald and Wolfowitz [11] proposed as test statistic the total number of runs of the ordered x's and y's, the hypothesis to be rejected if the number of runs is too small. The authors proved their test consistent, under the assumption of constant ratio of sample sizes m/n, against alternatives of all shapes restricted only by mild assumptions, concerning existence and positiveness of the probability densities. It was also proved in their paper that the test statistic has an asymptotically normal distribution when the hypothesis is true. More recently Wolfowitz [12] proved that the limiting dis-

tribution is normal even when $F \neq G$, and obtained the asymptotic variance for this case. It follows from his results that the test is in general not consistent if $m/n \to 0$ or ∞. This is actually what one would expect since when m/n is sufficiently extreme the maximum number of runs will in general occur with near-certainty whether the hypothesis is true or false.

Another test suitable for this problem is that of Smirnov [13] based on the maximum difference between the two sample cumulative distribution functions. For the given samples $X_1, \cdots, X_m$; $Y_1, \cdots, Y_n$ let

$$\phi_m(t) = \phi(X_1, \cdots, X_m; t) = \frac{1}{m} \text{ (number of } X\text{'s} \leqq t),$$

$$\psi_n(t) = \psi(Y_1, \cdots, Y_n; t) = \frac{1}{n} \text{ (number of } Y\text{'s} \leqq t),$$

be the two sample cumulative distribution functions. It follows from a theorem of Glivenko-Cantelli [14], that $\sup_t | \phi_m(t) - F(t) |$ and $\sup | \psi_n(t) - G(t) |$ tend to zero in probability as $\min(m, n) \to \infty$. From this it is easily seen that $\sup | \phi_m(t) - \psi_n(t) |$ is a consistent estimate of $\sup | F(t) - G(t) |$, and hence that Smirnov's test is consistent against all alternatives $F \neq G$ as $\min(m, n) \to \infty$. A different proof of this fact was given recently by Massey [25].

The large sample distribution of $\sup | \phi_m(t) - \psi_n(t) |$ was obtained by Smirnov, for the case that $F = G$, a simpler proof having recently been given by Feller [15] (see also Doob [16] and Smirnov [17]). Although the large sample distribution is not known when $F \neq G$, Massey [25] obtained a lower bound for the power of Smirnov's test, which may permit comparing this test with others.

While these two generally consistent tests are known for the two-sample problem, very little work has been done on the existence of unbiased tests for this or other nonparametric problems. Mann [18] proved unbiasedness of a test for randomness against a certain class of trends. Hoeffding [5] proved the non-existence for the hypothesis of independence of unbiased critical regions based on ranks, corresponding to certain very small levels of significance.

As far as the two-sample problem is concerned, Smirnov's test is easily shown to be biased on the basis of an example given by Massey for the problem of goodness of fit. On the other hand, it seems very possible that the Wald-Wolfowitz run test is unbiased whenever the two samples are of equal size. We have not proved this but shall now construct a test for the two sample problem that is strictly unbiased.

LEMMA 4.1 *Let X, X'; Y, Y' be independently drawn from populations with continuous cumulatives F, G respectively, and let us denote for any random variables U, U'; V, V' the event* $\max(U, U') < \min(V, V')$ *by $U, U' < V, V'$. Then*

$$p = P((X, X' < Y, Y') + (Y, Y' < X, X')) = \tfrac{1}{3} + 2 \int (F - G)^2 \, d\left(\frac{F + G}{2}\right),$$

and hence p attains its minimum value $\frac{1}{3}$ if and only if $F = G$.

Proof. Since F and G are continuous,

$$p = \int (1 - F)^2 \, dG^2 + \int (1 - G)^2 \, dF^2 = 2 + \int (F^2 \, dG^2 + G^2 \, dF^2)$$

$$- 4 \int FG \, d(F + G) = 2 + \int d(F^2G^2) - 4 \int FG \, d(F + G)$$

$$= 3 - 2 \int [(F+G)^2 - (F - G)^2] \, d\left(\frac{F + G}{2}\right)$$

$$= \tfrac{1}{3} + 2 \int (F - G)^2 \, d\left(\frac{F + G}{2}\right).$$

To prove the second part of the lemma, we must show that $\Delta = \int (F - G)^2 \, d(F + G) = 0$ implies $F = G$. Now $\Delta = 0$ implies $F(x) = G(x)$ except possibly on a set N such that $\int_N dF = \int_N dG = 1$. Suppose that $F(x_1) \neq G(x_1)$, $G(x_1) - F(x_1) = \eta > 0$ say. Then by continuity there exists $x_0 < x_1$ such that $G(x_0) = F(x_0) + \eta/2$ and $F(x) < G(x)$ for $x_0 \leqq x \leqq x_1$. Since $G(x_1) - G(x_0) > 0$, it follows that $\Delta > 0$.

It is now clear that there exists a strictly unbiased test of $H:F = G$ if $m, n \geqq 2$. For we can consider the number of quadruples $X_{2i-1}, X_{2i}; Y_{2i-1}, Y_{2i}$ for which either the two X's fall below the two Y's or vice versa. These may be regarded as the successes in independent trials with probability $p = \frac{1}{3} + 2\Delta$ of success, and the problem reduces to that of testing $H:p = \frac{1}{3}$ against alternatives $p > \frac{1}{3}$.

The unbiased test just described has the pleasant property that its power is a strictly increasing function of $\Delta = \int (F - G)^2 \, d\,\frac{F + G}{2}$, which seems a reasonable measure of the degree of difference of F and G. On the other hand one would not expect this test to be very efficient. More reasonable use of the data seems to be made if one modifies the test in the direction of the Mann-Whitney test described earlier. One would then compare each pair of X's with each pair of Y's, and reject H if among the $\binom{m}{2}\binom{n}{2}$ possible quadruples $X_i, X_j; Y_k, Y_l$ it happened too frequently that both X's lie on the same side of both Y's.

This test is no longer unbiased, but it is still consistent as follows from the argument given in the previous section. Further, the test retains the property that the statistic on which it is based provides an unbiased estimate, in fact the minimum variance unbiased estimate, of the quantity $\int (F - G)^2 \, d\,\frac{F + G}{2}$. Finally, it is again easily seen that the distribution of the test statistic is approximately normal, degeneracy occurring only if $P(Y > X)$ equals 1 or 0.

The test can be expressed in a form more convenient for computation in terms of the ranks of one of the sets of variables. Let $r_1 < r_2 < \cdots < r_n$ be the ranks

of the n Y's among the totality of $m + n$ observations, and denote by $Q_{m,n}$ the number of quadruples X_i, X_j, Y_k, Y_l for which both X's lie on the same side of both Y's. Then it is easily seen that

$$Q_{m,n} = \sum_{k=1}^{n} \left[(n - k) \binom{r_k - k}{2} + (k - 1) \binom{m - r_k + k}{2} \right].$$

From this it follows by easy computation that

$$2Q_{m,n} = (n - 1) \sum_{k=1}^{n} r_k^2 - 2(n + m - 2)\Sigma k r_k - (n - 2m + 1)\Sigma r_k$$
$$+ (n + 2m - 3) \frac{n(n + 1)(2n + 1)}{6} + (n + m^2 - 3m + 1) \frac{n(n + 1)}{2}.$$

It may perhaps be worth noting that the first of the two tests described in this section can also be used as the basis of a sequential test of the two sample problem. This is clear since the problem is simply reduced to that of testing a simple binomial situation against a one-sided class of alternatives. The sequential probability ratio test to which one is led in this manner of course is again unbiased and has a power function that is strictly increasing in $\int (F - G)^2 \, d \frac{F + G}{2}$.

The measure of discrepancy

$$\int (F - G)^2 \, d \frac{F + G}{2}$$

utilised in the tests of the present section, suggests using $\int (\phi_m - \psi_n) \, d \left(\frac{\phi_m + \psi_n}{2} \right)$ as test statistic. It should be pointed out that tests of this kind have been studied in connection with the closely related problem of goodness of fit by Cramér [19] and von Mises [20]. In the present case, let us denote the x's and y's in order of magnitude by $x^{(1)} < x^{(2)} < \cdots < x^{(m)}$; $y^{(1)} < y^{(2)} < \cdots < y^{(n)}$, let m_1 be the number of x's $< y^{(1)}$, m_2 the number of x's between $y^{(1)}$ and $y^{(2)}$ etc., and define $n_1, n_2, \cdots$ analogously. Then it is easily seen that

$$\int (\phi_m - \psi_n)^2 \, d(\phi_m + \psi_n)$$
$$= \frac{1}{n}\left[\left(\frac{m_1}{m} - \frac{1}{n} \right)^2 + \left(\frac{m_1 + m_2}{m} - \frac{2}{n} \right)^2 + \cdots + \left(\frac{m_1 + \cdots + m_n}{m} - 1 \right)^2 \right]$$
$$+ \frac{1}{m} \left[\left(\frac{n_1}{n} - \frac{1}{m} \right)^2 + \cdots \right].$$

Tests of this type have been proposed by Dixon and by Mood [21], but have not been studied thoroughly.

Finally it should be mentioned that one might also try the method of randomisation, which has been considered by Pitman [22] and others in connection

with specific classes of alternatives, for the present problem. One statistic which may be suitable for this purpose if $m = n$ is $\sum_{i=1}^{n} (Y^{(i)} - X^{(i)})^2$.

5. Discontinuous distributions. So far, we have assumed F and G to be continuous. This assumption is obviously not satisfied in practice, and we must therefore consider the difficulties introduced by discontinuities. (These difficulties were investigated in connection with various estimation problems by Scheffé and Tukey [23]).

Let us restrict our attention to rank tests and introduce the convention that tied observations are ordered at random. Thus if $X_{i_1} = \cdots = X_{i_s} = Y_{j_1} = \cdots = Y_{j_t}, s + t = r$, we perform an experiment with $r!$ possible and equally probable outcomes. We then establish a 1:1 correspondence between the $r!$ possible orderings of r objects and these $r!$ outcomes, and treat the X's and Y's as if they had occurred in the order indicated by this experiment. If the X's and Y's have the same distribution it is then clear that the distribution of any rank statistic of the X's and Y's is what it would be if this common distribution were continuous since in both cases each possible ordering of the X's and Y's is again equally probable.

In order to see that various unbiasedness results of the preceding and following sections remain valid, we state the following

LEMMA 5.1. *Let $\mathcal{E} = \mathcal{E}(X_1, \cdots, X_m, Y_1, \cdots, Y_n)$ be a random event depending only on the ranking of the X's and Y's. Suppose that F and G may have discontinuities and that in case of ties the event $\mathcal{E}$ is defined by ordering the tied observations at random. Then there exist continuous* c.d.f.'s *$F^* = F^*(F, G)$ and $G^* = G^*(F, G)$ such that*

$$P_{F,G}(\mathcal{E}) = P_{F^*,G^*}(\mathcal{E}).$$

and that $F^* = G^*$ if and only if $F = G$.

PROOF. We shall only give the construction of F^*, G^*; the remainder of the proof then follows easily.

Consider the (denumerable) totality of points that are points of discontinuity of either F or G, and suppose these points have been numbered: $x_1, x_2, \cdots$. Consider first x_1 and define two new c.d.f.'s F_1, G_1 as follows:

$$\begin{aligned} F_1(x) &= F(x + \tfrac{1}{4}) \quad \text{if} \quad x < x_1 - \tfrac{1}{4} \\ &= F(x_1^-) + \frac{x - (x_1 - \frac{1}{4})}{\frac{1}{2}} [F(x_1) - F(x_1^-)] \quad \text{if} \quad |x_1 - x| \leqq \tfrac{1}{4} \\ &= F(x - \tfrac{1}{4}) \quad \text{if} \quad x > x_1 + \tfrac{1}{4}. \end{aligned}$$

G_1 is defined analogously in terms of G. What this construction does is to push F and G apart at x_1 symmetrically by a total amount of $\frac{1}{2}$, and to distribute the probability at x_1 uniformly over the gap thus created.

In the same way we now push F_1 and G_1 apart at the second discontinuity (in its new position) by a total amount of $1/2^2$ and distribute the amount of jump

uniformly over the gap, thus obtaining F_2 and G_2. Then the sequence F_1, F_2, $\cdots$ will converge to a continuous distribution F^* and analogously for the G's and F^*, G^* will have the desired properties.

It follows from this lemma that the unbiased test of the hypothesis $F = G$ discussed in Section 4 remains strictly unbiased when the assumption of continuity of F and G is dropped. On the other hand, the power is no longer such a simple function of F and G. In fact let X, X'; Y, Y' denote as before independent random variables with distributions F and G respectively and denote by $X, X' < Y, Y'$ that this ordering occurred after randomisation of ties. Then it is not difficult to show that

$$P((X, X' < Y, Y') + (Y, Y' < X, X')) = \tfrac{1}{3} + 2\Delta',$$

where

$$3\Delta' = \int [(F - G)^2 + (F^- - G^-)^2 + (F - G)(F^- - G^-)]\, d\,\frac{F + G}{2}.$$

Here $F^-(x) = F(x^-)$, $G^-(x) = G(x^-)$.

6. Existence of unbiased tests for the hypothesis of independence and some other nonparametric problems. In this last section we shall briefly consider some more complicated nonparametric problems. Our aim is to prove for all these problems the existence of strictly unbiased and consistent tests. The problem is treated purely theoretically in that no effort is made to construct tests that make good use of the data and that are convenient to apply, but that instead the sole purpose is to exhibit tests possessing the properties asked for.

For the hypothesis of independence Hoeffding proposed a test that he proved consistent against all alternatives with continuous joint and marginal probability densities. In this connection he also considered the problem of unbiasedness and proved the nonexistence of unbiased critical regions based on rank for certain small levels of significance. This negative result seems to contradict those of the present section. This is however not so. Hoeffding restricted his attention to critical regions while we are here admitting also randomised tests. It should be pointed out in this connection that, while randomisation was used in previous sections only in a trivial manner, namely so as to get the exact level of significance, we shall here make very heavy use of this device. This could be avoided in part, however the tests would then become more complicated. Further if the problem is reduced, as is done here, to that of testing equality of two binomial p's, randomisation is needed to get an exactly similar test.

The hypothesis of independence states that the joint c.d.f. equals the product of the two marginal c.d.f.'s. Thus if $(X_i^{(1)}, X_i^{(2)})$, $i = 1, 2, \cdots$, are independently drawn from a bivariate distribution F, it is equivalent to the hypothesis that the pair $(X_1^{(1)}, X_1^{(2)})$ comes from the same bivariate population as the pair $(X_2^{(1)}, X_3^{(2)})$. It is therefore clear that if we can prove the existence of strictly unbiased and consistent tests for the bivariate two-sample problem, this will

imply the existence of tests with these properties for the hypothesis of independence. The same remark clearly applies to hypothesis of independence (both complete independence and independence of sets of variates) in more than two variables.

Consider now samples $X_i = (X_i^{(1)}, \cdots, X_i^{(k)})$ $i = 1, 2, \cdots$ and $Y_j = (Y_j^{(1)}, \cdots, Y_j^{(k)})$ $j = 1, 2, \cdots$ from two k-variate distributions F and G. The work of section 4 suggests utilising the expression

$$\int (F - G)^2 \, d\left(\frac{F + G}{2}\right) = \int (F^2 + G^2) \, d\left(\frac{F + G}{2}\right) - 2\int FG \, d\left(\frac{F + G}{2}\right).$$

All that is necessary is to construct events A and B such that

$$p_1 = P(A) = \int \frac{F^2 + G^2}{2} \, d\left(\frac{F + G}{2}\right),$$

$$p_2 = P(B) = \int FG \, d\left(\frac{F + G}{2}\right).$$

The hypothesis $H:F = G$ will then be reduced to $H':p_1 = p_2$ to be tested against alternatives $p_1 > p_2$. The events A and B may be defined as follows:

A: With probability $\frac{1}{2}$ observe either X_1, X_2 or Y_1, Y_2 and with probability $\frac{1}{2}$ observe either X_3 or Y_3. Denote the three variables that are observed by Z_1, Z_2, Z_3, and define A as the event

$$Z_1^{(i)}, Z_2^{(i)} \leqq Z_3^{(i)} \qquad \text{for} \quad i = 1, \cdots, k.$$

B: Observe X_4, Y_4 and with probability $\frac{1}{2}$ either X_5 or Y_5. If the last of these variables is denoted by Z_5, define B as the event

$$X_4^{(i)}, Y_4^{(i)} \leqq Z_5^{(i)} \qquad \text{for} \quad i = 1, \cdots, k.$$

It should be mentioned that instead of observing five random vectors some of which may be either X's or Y's, we could have obtained a test with the desired property based on ten observations, five X's and five Y's.

To complete the proof we must show that the hypotheses H and H' are really equivalent, that is, that $p_1 = p_2$ if and only if $F = G$. For the case that F and G are continuous this follows immediately by an argument similar to the one given in the univariate case, and it is easy to show it even without this restriction.

It is clear that one can generalise further and instead of two samples consider s samples. For this purpose one may replace $\int (F - G)^2 \, d\left(\frac{F + G}{2}\right)$ for example by $\sum_{i=1}^{s} (F_i - \bar{F})^2 \, d\bar{F}$ where $\bar{F}$ is the average of the s c.d.f.'s. Alternatively, one may utilise the expression $\sum_{i<j} \int (F_i - F_j)^2 \, d\left(\frac{F_i + F_j}{2}\right)$.

As a last problem let us consider a sample $X_1, \cdots, X_n$ from an unknown

univariate c.d.f. F, assumed to be continuous. It is desired to test the hypothesis H of symmetry with respect to the origin, i.e., that $F(x) = 1 - F(-x)$ for all x. Smirnov [24] recently proposed $\max_x \{ | N^+(x) - N^-(x) | \}$ as a test statistic where $N^+(x)$, $N^-(x)$ denote the number of x's contained in the intervals $(0, x)$, $(-x, 0)$ respectively.

The work of Section 4 suggests considering 4 X's (X_i, X_j, X_k, X_l) and defining the following two events.

A: Exactly two of the four X's are positive.

B: If A is satisfied, and $X_i, X_j < 0 < X_k, X_l$, say, the event B is said to occur if neither

$$|X_i|, |X_j| < X_k, X_l \quad \text{nor} \quad X_k, X_l < |X_i|, |X_j|.$$

Then if $F(0) = p_0$, $P(A) = 6p_0^2q_0^2$ takes on its maximum value 3/8 if and only if $p_0 = 1/2$. Further, $P(B \mid A)$ takes on its maximum value 2/3 if and only if the conditional distribution F^* of $-X$ given $X < 0$ is the same as that, G^*, of X given $X > 0$. Thus

$$P(AB) = 6p_0^2q_0^2\left\{\frac{2}{3} - 2\int (F^* - G^*)^2 \, d\left(\frac{F^* + G^*}{2}\right)\right\}$$

takes on its maximum value 1/4 if and only if the hypothesis of symmetry holds.

If we apply this method to independent quadruples, we obtain a test that is strictly unbiased and consistent. If we apply it to all possible quadruples the test remains consistent and may be a reasonable test for the hypothesis in question. Hoeffding's theory can again be applied to the asymptotic distribution problem.

REFERENCES

[1] H. C. Mathisen, "A method of testing the hypothesis that two samples are from the same population," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 188–194.

[2] W. Hoeffding, "A class of statistics with asymptotically normal distributions," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 293–325.

[3] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 50–60.

[4] W. R. Thompson, "Biological applications of normal range and associated significance tests in ignorance of original distribution forms," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 281–287.

[5] W. Hoeffding, "A non-parametric test of independence," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 546–557.

[6] P. R. Halmos, "The theory of unbiased estimation," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 34–43.

[7] E. L. Lehmann and H. Scheffé, "Completeness, similar regions, and unbiased estimation—Part I," *Sankhyā*, Vol. 10 (1950), pp. 305–340.

[8] E. L. Lehmann and H. Scheffé, "Completeness, similar regions, and unbiased estimation, Part II," unpublished.

[9] J. Berkson, "Comments on Dr. Madow's 'Note on tests of departure from normality' with some remarks concerning tests of significance," *Jour. Am. Stat. Assn.*, Vol. 36 (1941), p. 539.
[10] H. Cramér, "Contributions to the theory of statistical estimation," *Skandinavisk Aktuarietidskrift*, Vol. 29 (1946), p. 85.
[11] A. Wald and J. Wolfowitz, "On a test whether two samples are from the same population," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 147–162.
[12] J. Wolfowitz, "Non-parametric statistical inference," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1949.
[13] N. V. Smirnov, "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," *Bull. Math. Univ. Moscow, Serie Int.*, Vol. 2 (1939).
[14] M. Fréchet, *Recherches Théoriques Modernes sur la Théorie des Probabilités, Vol. 1*, p. 260, Gauthier-Villars, Paris, 1937.
[15] W. Feller, "On the Kolmogorov-Smirnov limit theorems for empirical distributions," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 177–189.
[16] J. L. Doob, "Heuristic approach to the Kolmogorov-Smirnov theorems," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 393–403.
[17] N. V. Smirnov, "Approximate laws of distribution of random variables from empirical data," *Uspehi Matem. Nauk*, Vol. 10 (1944), p. 179.
[18] H. B. Mann, "Non-parametric tests against trend," *Econometrica*, Vol. 13 (1945), pp. 245–259.
[19] H. Cramér, "On the composition of elementary errors," *Skandinavisk Aktuarietidskrift*, Vol. 11 (1928), p. 13 and p. 141.
[20] R. von Mises, *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*, Franz Deuticke, Leipzig, 1931.
[21] W. Dixon, "A criterion for testing the hypothesis that two samples are from the same population," *Annals of Math Stat.*, Vol. 11 (1940), pp. 199–204.
[22] E. J. G. Pitman, "Significance tests which may be applied to samples from any populations," *Jour. Roy. Stat. Soc., Suppl.*, Vol. 4 (1937), pp. 225–232.
[23] H. Scheffé and J. W. Tukey, "Non-parametric estimation. I. Validation of order statistics," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 187–192.
[24] N. V. Smirnov, "Sur un critère de symetrie de la loi de distribution d'une variable aléatoire," *C. R. Acad. Sci. URSS*, Vol. 56 (1947), p. 11.
[25] F. J. Massey, Jr., "A note on the power of a non-parametric test," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 440–443.
[26] H. R. Van der Vaart, "Some remarks on the power function of Wilcoxon's test for the problem of two samples. I," *Indagationes Math.*, Vol. 12 (1950), pp. 146–158.

A SIGNIFICANCE TEST FOR EXPONENTIAL REGRESSION

By E. S. Keeping

University of Alberta[1]

1. Summary. A general method of testing the significance of nonlinear regression, suggested by Hotelling, is adapted to the regression equations $Y = be^{px}$ and $Y = a + be^{px}$. The values of x are taken to be in arithmetic progression, and the standard deviation of the observed y is supposed constant for all x. This is in contrast to the assumption, implicit in the usual procedure of fitting a straight line to $\log y$, that the standard deviation of $\log y$ is constant.

It will be observed that the distribution of $y_1, y_2, \cdots, y_n$ must be such that the joint probability density for $y_1, y_2, \cdots, y_n$ is a function of $x_1^2 + x_2^2 + \cdots + x_n^2$, and this condition implies the assumption of normality. The null hypothesis is that $be^{px} = 0$ for all x, while the alternative hypotheses are specified by $b \neq 0, p \neq -\infty$.

The method involves the calculation of the volume of a "tube" on a hypersphere in n-dimensional space. An asymptotic expression for the length of the tube is developed, and it is shown that the curvature of the axis is everywhere finite. From this expression, for values of the correlation coefficient R between observed and fitted values of y at least as great as 0.894, a function of R is obtained giving the probability that a random sample would yield at least as great a value of R.

A short table giving R for various significance levels and various sizes of sample is calculated for each of the equations mentioned, and the application to certain experimental data is discussed.

2. Introduction. Some years ago, Hotelling [1] suggested a geometrical method of determining the significance of the correlation coefficient corresponding to a fitted regression of y upon x, when y is a random variable and the values of x are known. Suppose that a curve of the form

$$Y = bf(x, p), \tag{2.1}$$

where b, p are constants and $f(x, p)$ is not identically zero, is fitted to a set of observations $y_1, y_2, \cdots, y_n$, which are assumed to be independently and normally distributed about zero, with the same variance σ^2. The null hypothesis is that $b = 0$, while the alternative hypothesis is that b is not zero. By the principle of least squares, we minimize

$$\Sigma (y_\alpha - Y_\alpha)^2 = \Sigma [y_\alpha - bf(x_\alpha, p)]^2. \tag{2.2}$$

The set of values $y_1, \cdots, y_n$ defines a point in n-dimensional space. The set

[1] Research carried out at the Institute of Statistics, University of North Carolina.

$Y_1, \cdots, Y_n$ also defines a point which lies on the 2-dimensional hypersurface defined parametrically in terms of b, p by the n equations:

$$Y_\alpha = bf(x_\alpha, p), \qquad \alpha = 1, 2, \cdots, n. \tag{2.3}$$

If θ is the angle between the lines joining the origin to $(y_1, \cdots, y_n)$ and $(Y_1, \cdots, Y_n)$,

$$\cos\theta = \Sigma\, y_\alpha Y_\alpha / [\Sigma\, y_\alpha^2\, \Sigma\, Y_\alpha^2]^{\frac{1}{2}}, \tag{2.4}$$

and this is the correlation coefficient R between the observed and fitted values, calculated without elimination of the mean. The least squares process is thus equivalent to maximizing R, or minimizing θ.

Since by the null hypothesis the point $(y_1, \cdots, y_n)$ lies with uniform probability density anywhere on the surface of a sphere whose centre is the origin, the density function of the projection of the point $(y_1, \cdots, y_n)$ on the unit hypersphere has complete spherical symmetry around the origin. This will also be true if the joint probability density for $y_1, y_2, \cdots, y_n$ is any function of $x_1^2 + x_2^2 + \cdots + x_n^2$, and so is constant on the hypersphere $x_1^2 + \cdots + x_n^2 = 1$. The probability that R is greater than some fixed value R_0 is therefore, for a given Y, proportional to the "volume" of the sphere, in the $(n-1)$-dimensional spherical space, having centre Y and geodesic radius $\theta_0 = \cos^{-1} R_0$. The total probability that R lies between R_0 and 1 is therefore given by the ratio of the "volume" of the "tube" of geodesic radius θ_0, surrounding the curve formed by the projection of Y, to the total "area" of the unit hypersphere.

Hotelling [1] has shown that the volume of such a tube on a hypersphere in n-dimensional space is equal to the length of the curve multiplied by

$$\pi^{\frac{1}{2}(n-2)} \sin^{n-2} \theta_0 / \Gamma(n/2),$$

provided that the curve is closed, and nowhere has a radius of geodesic curvature less than $\sin\theta_0$, and provided also that portions of the tube corresponding to nonconsecutive arcs of the axial curve do not overlap. If the curve has ends, there will be hemispherical "caps" at the ends of the tube to be added to the total volume.

This geometrical method was applied by D. M. Starkey [2] to the case of periodogram analysis, in which there are additional parameters, so that the projection of Y is not a curve but a surface. The practical application of the method in this case is limited to quite small values of θ_0, because of the approximations necessary in the evaluation of the integrals involved.

The 3-parameter equation

$$Y_\alpha = a + bf(x_\alpha, p) \tag{2.5}$$

is readily reducible, theoretically, to the form treated above. Minimizing $\Sigma\, [y_\alpha - a - bf(x_\alpha, p)]^2$ is equivalent to minimizing $\Sigma\, (y'_\alpha - Y'_\alpha)^2$, where y'_α, Y'_α are the projections of y_α, Y_α on the hyperplane $\Sigma\, y_\alpha = 0$. Since $y'_\alpha = y_\alpha - \bar{y}$ and $Y'_\alpha = b(f_\alpha - \bar{f})$, where f_α stands for $f(x_\alpha, p)$, the angle θ between

the lines joining the origin to $(y'_1, y'_2, \cdots, y'_n)$ and $(Y'_1, Y'_2, \cdots, Y'_n)$ is given by

$$\cos\theta = \Sigma (y_\alpha - \bar{y})(Y_\alpha - \bar{Y})/[\Sigma (y_\alpha - \bar{y})^2 \Sigma (Y_\alpha - \bar{Y})^2]^{\frac{1}{2}} \tag{2.6}$$

and so is equal to the correlation coefficient R between observed and fitted values, calculated in the usual way with elimination of the means. The point $(Y'_1, Y'_2, \cdots, Y'_n)$ lies on a 2-dimensional projection of the 3-dimensional hypersurface defined parametrically by (2.5). If we now project from the origin on to a hypersphere of $n - 2$ dimensions (intrinsically) in the hyperplane $\Sigma y_\alpha = 0$, the projection of $(Y'_1, Y'_2, \cdots, Y'_n)$ will be a point $(Y''_1, Y''_2, \cdots, Y''_n)$ lying on a curve on the surface of this hypersphere. The method already given therefore applies in this case, with the appropriate change in the dimensionality of the hypersphere.

The present paper deals with the application to exponential regression. The curves to be fitted are

$$Y = be^{px} \tag{2.7}$$

and

$$Y = a + be^{px}, \tag{2.8}$$

the latter of which will be referred to as the "modified exponential equation." The mathematical difficulties are increased greatly by the additional constant.

3. Formulas for projections of regression curves. We suppose that the fixed values of x are equidistant, and choose units so that

$$x_1 = 1, \quad x_2 = 2, \quad \cdots, \quad x_n = n. \tag{3.1}$$

The corresponding values of Y are

$$Y_1 = bq, \quad Y_2 = bq^2, \quad \cdots, \quad Y_n = bq^n, \tag{3.2}$$

where $q = e^p$. If the projections on the unit hypersphere are denoted by $Y'_1, \cdots, Y'_n$,

$$Y'_\alpha = \lambda q^\alpha, \quad \Sigma Y'^2_\alpha = 1. \tag{3.3}$$

Hence

$$\lambda^2 = q^{-2}(1 - q^2)(1 - q^{2n})^{-1} \tag{3.4}$$

and

$$Y'_\alpha = q^{\alpha-1}\left(\frac{1 - q^2}{1 - q^{2n}}\right)^{\frac{1}{2}}. \tag{3.5}$$

The element ds of the curve formed by Y' $(0 \leq q < \infty)$ is given by

$$\begin{aligned}(ds)^2 &= \Sigma(dY'_\alpha)^2 \\ &= \frac{1 - q^2}{1 - q^{2n}} \sum_{\alpha=1}^{n} \left\{\alpha - 1 + \frac{nq^{2n}}{1 - q^{2n}} - \frac{q^2}{1 - q^2}\right\}^2 q^{2\alpha-4}(dq)^2.\end{aligned} \tag{3.6}$$

This reduces, after some algebraic manipulation, to

$$\frac{ds}{dx} = (2x)^{-1}\left[\frac{x}{(1-x)^2} - \frac{n^2x^n}{(1-x^n)^2}\right]^{\frac{1}{2}}, \tag{3.7}$$

where $x = q^2$. The length of the projected curve is obtained by integrating from 0 to ∞.

For the modified exponential equation, we have, instead of (3.3),

$$Y''_{\alpha} = \lambda(q^{\alpha} - f/n), \tag{3.8}$$

where

$$f = \Sigma\, q^{\alpha} = q(1 - q^n)(1 - q)^{-1}. \tag{3.9}$$

Since $\Sigma\,(Y''_{\alpha})^2 = 1$, we obtain

$$\lambda = (g - f^2/n)^{-\frac{1}{2}},$$

where

$$g = q^2(1 - q^{2n})/(1 - q^2). \tag{3.10}$$

The expression for $(ds/dq)^2 = \Sigma\,(dY''_{\alpha}/dq)^2$ reduces after lengthy algebra to

$$\left(\frac{ds}{dq}\right)^2 = \frac{1}{(1-q^2)^2} - \left\{\frac{\dfrac{nq^{n-1}}{1-q^n} - \dfrac{1}{n}\cdot\dfrac{1-q^n}{(1-q)^2}}{1 + q^n - \dfrac{1}{n}\cdot\dfrac{1+q}{1-q}(1-q^n)}\right\}^2, \tag{3.11}$$

whence s may be obtained by integration.

4. Lengths of the projected curves. From (3.7)

$$\begin{aligned} l_n &= \int_0^{\infty} [x(1-x)^{-2} - n^2x^n(1-x^n)^{-2}]^{\frac{1}{2}}\,dx/(2x) \\ &= \int_0^1 [x(1-x)^{-2} - n^2x^n(1-x^n)^{-2}]^{\frac{1}{2}}\,dx/x, \end{aligned} \tag{4.1}$$

since the substitution $x = 1/y$ leaves the integral unchanged.

For $n = 2$ the integral is elementary and reduces to $\pi/2$. For $n = 3$ it can be expressed in terms of elliptic integrals of the first and third kinds. For higher values of n a convergent series can be obtained, which, however, converges very slowly for n larger than 5.

Putting $x = (1 - u)/(1 + u)$, $g = 1 - u^2$, we obtain

$$\begin{aligned} l_n &= 2\int_0^1 \left[\frac{1-u^2}{4u^2} - \frac{n^2(1-u^2)^n}{\{(1+u)^n - (1-u)^n\}^2}\right]^{\frac{1}{2}} \frac{du}{1-u^2} \\ &= 2\int_0^1 g^{-\frac{1}{2}}\left[\frac{1}{4(1-g)} - \frac{n^2g^{n-1}}{\sum_{t=0}^{2n}\binom{2n}{t}u^t - 2g^n + \sum_0^{2n}\binom{2n}{t}(-u)^t}\right]^{\frac{1}{2}} du, \end{aligned} \tag{4.2}$$

and this integral may be shown to exist for every finite n. For $n = 3$, we have

$$l_3 = \pi/2\,[1 + 1/4 + 9/256 + 5/512 + 385/262144 + \cdots]$$
$$= 2.037.$$

For $n = 4$, a similar method gives $l_4 = 2.35$, and for $n = 5$ we obtain $l_5 = 2.58$. However, as n increases, the method is more laborious and the integrand does not converge so rapidly, so that an asymptotic expression for l_n is more convenient.

For the case $n = 3$, (3.11) reduces to

$$\frac{ds}{dq} = \frac{\sqrt{3}}{2} \cdot \frac{1}{1 + q + q^2},$$

whence $l_3' = \pi/3$. That this is correct may be seen by visualizing the projection of the regression curve on to a circle in the plane through the origin which is equally inclined to all three axes. Writing $u = \tanh p/2$, we obtain

(4.3) $$l_n' = \int_0^1 [1 - \varphi^2]^{\frac{1}{2}}\, du/u,$$

where

(4.4) $$\varphi = \left[1 - \frac{4n^2u^2(1 - u^2)^{n-1}}{\{(1 + u)^n - (1 - u)^n\}^2}\right] \div \left[\frac{nu\{(1 + u)^n + 1 - u)^n\}}{(1 + u)^n - (1 - u)^n} - 1\right].$$

It may be shown that the denominator of φ never vanishes and that $\varphi \leq 1$ in the region of integration.

When $n = 4$, the integral (4.3) is elliptic, and we find $l_4' = 1.418$.

$$\text{When} \quad n = 5, \qquad \varphi = \frac{25 - 5u^2 + 15u^4 - 3u^6}{25 + 65u^2 + 35u^4 + 3u^6}.$$

The integral may be evaluated by quadrature. The numerical value is $l_5' = 1.675$. For larger values of n, an approximation is obtained in the next section.

5. Approximations to the length. Putting $x = e^{-2v}$, we have from (4.1)

(5.1) $$l_n = \int_0^\infty \left(\frac{1}{\sinh^2 v} - \frac{n^2}{\sinh^2 nv}\right)^{\frac{1}{2}} dv.$$

For values of v in the range 0 to $1/n$ we may write the integrand as a series and integrate term by term. Thus, if

$$I_1 = \int_0^{1/n} \left(\frac{1}{\sinh^2 v} - \frac{n^2}{\sinh^2 nv}\right)^{\frac{1}{2}} dv,$$

we obtain

$$I_1 = 0.559 - 0.298n^{-2} - 0.0594n^{-4} - \cdots. \tag{5.2}$$

Let

$$\begin{aligned} I_2 &= \int_{1/n}^{\infty} \left(\frac{1}{\sinh^2 v} - \frac{n^2}{\sinh^2 nv}\right)^{\frac{1}{2}} dv \\ &= 2\int_0^{e^{-1/n}} \frac{1}{1-u^2}\left[1 - \frac{n^2u^{2n-2}(1-u^2)^2}{(1-u^{2n})^2}\right]^{\frac{1}{2}} du. \end{aligned} \tag{5.3}$$

The second term in the square bracket is less than 1 at both ends of the range. Also, it has no maximum within the range. Hence the bracket may be expanded and integrated term by term, giving

$$\begin{aligned} I_2 &= 2\int_0^{e^{-1/n}} \frac{1}{1-u^2}\left[1 - \frac{n^2(1-u^2)^2u^{2n-2}}{2(1-u^{2n})^2} - \frac{n^4(1-u^2)^4u^{4n-4}}{8(1-u^{2n})^4} - \cdots\right] du \\ &= I_{21} + I_{22} + I_{23} + \cdots, \end{aligned}$$

where

$$I_{21} = 2\int_0^{e^{-1/n}} \frac{du}{1-u^2} = \log 2n + \frac{1}{12n^2} - \frac{1}{30n^3} + O\left(\frac{1}{n^4}\right). \tag{5.4}$$

Also,

$$\begin{aligned} I_{22} &= -n^2\int_0^{e^{-1/n}} u^{2n-2}(1-u^2)(1-u^{2n})^{-2}\, du \\ &= -n^2\int_0^{e^{-1/n}} [(u^{2n-2} - u^{2n}) + 2(u^{4n-2} - u^{4n}) + 3(u^{6n-2} - u^{6n}) + \cdots]\, du, \end{aligned}$$

which, on integrating, expanding the exponentials, and collecting terms, becomes

$$\begin{aligned} I_{22} = &-[(3/2)e^{-2} + (5/4)e^{-4} + (7/6)e^{-6} + (9/8)e^{-8} + \cdots] \\ &- 1/n^2[(19/24)e^{-2} + (71/192)e^{-4} + (61/216)e^{-6} + (379/1536)e^{-8} \\ &+ \cdots] + O(n^{-4}) \\ = &-0.229 - 0.114n^{-2} + O(n^{-4}). \end{aligned} \tag{5.5}$$

Similarly,

$$\begin{aligned} I_{23} &= -\frac{n^4}{4}\int_0^{e^{-1/n}} u^{4n-4}(1-u^2)^3(1-u^{2n})^{-4} du \\ &= -0.029 - 0.029n^{-2} + O(n^{-4}). \end{aligned} \tag{5.6}$$

Later terms in I_2 can be computed in a similar way, but the numerical factors diminish rapidly. Collecting terms from (5.2), (5.4), (5.5), (5.6), we get finally

$$l_n = \log n + 0.990 - 0.358n^{-2} + O(n^{-4}). \tag{5.7}$$

As an indication of the accuracy of this approximation, the value of l_n, neglecting terms of order n^{-4}, has been calculated in Table I for several values of n. It is not, of course, to be expected that the approximation will be very good for small n, although it is actually quite close even for $n = 3$ and $n = 4$.

TABLE I

Length of axis of tube ($Y = be^{pz}$)

n	Asymptotic value of l_n	Exact value of l_n
2	1.59	1.57
3	2.05	2.04
4	2.35	2.34
5	2.59	2.59
6	2.77	
7	2.93	
8	3.06	
9	3.18	
10	3.29	
12	3.47	
15	3.70	
20	3.98	
50	4.90	
100	5.60	

For the modified exponential equation, the above method is apparently not practicable with the more complicated integral (4.3). However, it is possible to obtain an approximate expression which will be valid for large n, although the agreement with the numerical values for $n = 3$, 4, and 5 is not very close.

In terms of $v = p/2$, (4.3) may be written

$$l'_n = \int_0^\infty 2(\sinh 2v)^{-1} \left[1 - \left\{\frac{1 - n^2 \sinh^2 v(\sinh nv)^{-2}}{n \tanh v(\tanh nv)^{-1} - 1}\right\}^2\right]^{\frac{1}{2}} dv. \tag{5.8}$$

For values of v between 0 and $1/n$, the integrand may be expanded in powers of v and integrated term by term. The result is

$$I_1 = 0.500 - 1.08n^{-2} + O(n^{-4}). \tag{5.9}$$

For any fixed u between 0 and 1, the function φ in (4.4) tends to $(nu - 1)^{-1}$ as $n \to \infty$, so that the integrand in (4.3) tends to $u^{-1}[1 - (nu - 1)^{-2}]^{\frac{1}{2}}$. However, this approximation is clearly not useful for u near $1/n$. But if we put $u = k/n$, where k is a fixed integer, then, for large n, φ tends to the value

$$\varphi = \frac{e^k - e^{-k} - 4k^2(e^k - e^{-k})^{-1}}{(k - 1)e^k + (k + 1)^{-k}}. \tag{5.10}$$

For fairly large k, this is very close to $(k - 1)^{-1}$. Thus for $k = 7$, it is 0.1667. Hence from $u = 7/n$ to 1 we can approximate φ asymptotically by $(nu - 1)^{-1}$, and obtain

$$I_2 \sim \int_{7/n}^{1} [1 - (nu - 1)^{-2}]^{\frac{1}{2}} u^{-1} \, du = \int_{w_1}^{w_2} [1 - \operatorname{sech} w] \, dw,$$

where $nu - 1 = \cosh w$ and w_1, w_2 are the values of w corresponding to $u = 7/n$ and $u = 1$ respectively. Hence

$$\begin{aligned} I_2 &\sim w_2 - w_1 - \tan^{-1} \sinh w_2 + \tan^{-1} \sinh w_1 \\ &= -1.952 + \log n + (1/4)n^{-2} + O(n^{-3}) \quad (n \geq 7). \end{aligned} \tag{5.11}$$

It remains to integrate between $\tanh(1/n)$ and $7/n$. From $1/n$ to $7/n$, an approximation may be obtained by quadrature, the ordinates being calculated from (5.10) for values of k between 1 and 7 inclusive. This gives, by Simpson's rule, a value 1.573. A small correction may be made for the integral between $\tanh(1/n)$ and $1/n$. Since $\tanh(1/n) = 1/n - (1/3)n^{-3}$, approximately, and since for $u = 1/n$, $u^{-1}[1 - \varphi^2]^{\frac{1}{2}} = 0.475n$ approximately, this integral will be $0.158n^{-2}$, neglecting terms of higher order.

Hence the final expression for the length of the axis of the tube is

$$l'_n = \log n + 0.121 - 0.67n^{-2} + O(n^{-3}). \tag{5.12}$$

Table II gives a few numerical values of l'_n, neglecting $O(n^{-3})$.

TABLE II

Length of axis of tube ($Y = a + be^{px}$)

n	Asymptotic value of l'_n	Correct value of l'_n
3	1.14	1.05
4	1.47	1.42
5	1.70	1.68
6	1.89	
7	2.05	
8	2.19	
9	2.31	
10	2.42	
15	2.83	
20	3.12	

6. Curvature of the projected curve. It was shown by Hotelling [1] that to avoid difficulties connected with local overlapping, or "kinking", of the tube surrounding the projected curve, it is necessary and sufficient that $\sin \theta \leq \rho$,

where θ is the geodesic radius of the tube and ρ the radius of geodesic curvature of its axis. In this section we show that the radius of curvature is always finite and greater than $1/\sqrt{5} = 0.447$. The statement by Hotelling (*loc. cit.*, p. 452), that the radius of curvature of the projected curve corresponding to $Y = be^{px}$ becomes zero at $p = \pm\infty$, appears to be in error.

The radius of curvature ρ with which we are dealing is defined by

$$\begin{aligned}\rho^{-2} &= \Sigma_\alpha(d^2Y'_\alpha/\,ds^2)^2\\ &= (ds/dp)^{-6}[(ds/dp)^2\Sigma(d^2Y'_\alpha/dp^2)^2 + (d^2s/dp^2)^2\Sigma(dY'_\alpha/dp)^2\\ &\qquad - 2ds/dp\cdot d^2s/dp^2\Sigma(dY'_\alpha/dp\cdot d^2Y'_\alpha/dp^2)].\end{aligned} \tag{6.1}$$

Since

$$\Sigma(dY'_\alpha/dp)^2 = (ds/dp)^2,\ \Sigma(dY'_\alpha/dp\cdot d^2Y'_\alpha/dp^2) = ds/dp\cdot d^2s/dp^2,$$

this reduces to

$$\rho^{-2} = (ds/dp)^{-4}[\Sigma(d^2Y'_\alpha/dp^2)^2 - (d^2s/dp^2)^2]. \tag{6.2}$$

In the present problem, $Y'_\alpha = \lambda e^{\alpha p}$, so that

$$d^2Y'_\alpha/dp^2 = e^{\alpha p}(d^2\lambda/dp^2 + 2\alpha\, d\lambda/dp + \alpha^2\lambda),$$

where

$$\lambda^{-2} = \Sigma e^{2\alpha p} = e^{2p}(e^{2np} - 1)(e^{2p} - 1)^{-1}.$$

After some reduction we obtain

$$\begin{aligned}\Sigma(d^2\,Y'_\alpha/dp^2)^2 &= \frac{3}{16}\left[\frac{1}{\sinh^2 p} - \frac{n^2}{\sinh^2 np}\right]^2\\ &\qquad + \frac{1}{8}\left[\frac{3 + 2\sinh^2 p}{\sinh^4 p} - n^4\,\frac{3 + 2\sinh^2 np}{\sinh^4 np}\right].\end{aligned} \tag{6.3}$$

From equation (3.7) we have, in terms of $p = \log q$ as parameter,

$$ds/dp = \tfrac{1}{4}[(\sinh p)^{-2} - n^2(\sinh np)^{-2}], \tag{6.4}$$

whence

$$d^2s/dp^2\cdot ds/dp = \tfrac{1}{4}[-\cosh p(\sinh p)^{-3} + n^3\cosh np(\sinh np)^{-3}]. \tag{6.5}$$

Therefore, from (6.2),

$$\begin{aligned}\rho^{-2} &= 3 + \left[\frac{6 + 4\sinh^2 p}{\sinh^4 p} - n^4\,\frac{6 + 4\sinh^2 np}{\sinh^2 np}\right]\left[\frac{1}{\sinh^2 p} - \frac{n^2}{\sinh^2 np}\right]^{-2}\\ &\qquad - 4\left[\frac{\cosh p}{\sinh^3 p} - \frac{n^3\cosh np}{\sinh^3 np}\right]^2\left[\frac{1}{\sinh^2 p} - \frac{n^2}{\sinh^2 np}\right]^{-3}.\end{aligned} \tag{6.6}$$

Now as $p \to 0$, the right hand side tends to $3 - [6(n^2 + 1)]/[5(n^2 - 1)]$. For $n = 2$ this is 1, as it should be, and as $n \to \infty$ it approaches the value 9/5. It

may be shown also that for $n \geq 4$, $\rho^{-2} \to 5 - 2e^{-2|p|}$ as $p \to \pm\infty$, so that $1/\rho^2 = 5$ for moderate values of p.

If $u = (\sinh^2 p)^{-1} - n^2 (\sinh^2 np)^{-1}$, the expression for ρ^{-2} can be written as

$$\rho^{-2} = 3 - u'^2u^{-3} + u''u^{-2}. \tag{6.7}$$

Hence

$$d(\rho^{-2})/dp = u^{-4}[3u'^3 - 4uu'u'' + u^2u''']. \tag{6.8}$$

By expanding in powers of e^{-p} it can be shown that the terms in e^{2p} and the constant term vanish, so that

$$d(\rho^{-2})/dp = O(e^{-2p}), \qquad p > 0,$$

for $n \geq 4$. Hence ρ^{-2} has no maximum or minimum at any point p, apart from the minimum at $p = 0$. The radius of curvature is therefore finite and remains between $1/\sqrt{5}$ and $[(5n^2 - 5)/(9n^2 - 21)]^{\frac{1}{2}}$, i.e., between 0.447 and 0.745, for any $n \geq 4$.

The condition for no local overlapping at any point of the tube is, therefore, $\sin\theta \leq 0.447$, or equivalently, $\cos\theta \geq 0.894$, where θ is the geodesic radius of the tube.

For the modified exponential curve, (6.2) still holds, with Y'_α replaced by $Y''_\alpha = (g - f^2/n)^{-\frac{1}{2}} (e^{\alpha p} - f/n)$. We now have

$$(ds/dp)^2 = (4\sinh^2 p)^{-1}\left[1 - \frac{1 - n^2\sinh^2(p/2)(\sinh np/2)^{-2}}{n\tanh(p/2)(\tanh np/2)^{-1} - 1}\right]. \tag{6.9}$$

I have not been able to obtain an explicit expression for the curvature of the axis of the tube, similar to (6.6). However, for small values of p, Y''_α and ds/dp may be expressed in series of powers of p, and I find after much algebraic calculation that when $p \to 0$,

$$1/\rho^2 \to \frac{19n^4 - 212n^2 + 544}{7n^4 - 56n^2 + 112}. \tag{6.10}$$

For $n = 3$ this reduces to 1, as it should, since in this case the curve is an arc of a unit circle.

As $n \to \infty$, $\rho \to (7/19)^{\frac{1}{2}} = 0.607$, so that the radius of curvature at the centre of the axial curve of the tube lies between 0.607 and 1 for all values of $n > 3$.

To find the curvature at the ends of the axial curve we need the limit of $1/\rho$ as $p \to \pm\infty$. Since the curve is symmetrical about $p = 0$, it is sufficient to consider $p \to \infty$. For $n \geq 5$, it may be shown that, as $p \to \infty$,

$$1/\rho^2 \to \frac{5n^5 - 35n^4 + 78n^3 - 108n - 12}{n(n-1)(n-2)^3}. \tag{6.11}$$

As $n \to \infty$, $\rho \to 1/\sqrt{5} = 0.446$, as for the simpler case treated above. For $n = 6$, ρ has the value 0.519.

Intuitively, one would expect for $Y = a + be^{px}$, as we have shown for $Y = be^{px}$, that the curvature of the projection would vary monotonically between the centre and the ends. A proof of this would, however, be desirable. Assuming that this statement is true, we have as the condition for no local overlapping that $\cos\theta \geq 0.894$.

It is of interest to see how much of the length of the tube will have a radius of curvature near to the minimum value 0.477, which corresponds to $p = \pm\infty$. Now, for $n \geq 5$, and for large p, equation (6.6) can be written

$$\rho^{-2} = 5 - 2e^{-2|p|} + O(e^{-6|p|}), \tag{6.12}$$

so that for $|p| > 1$, the first two terms may be considered a fair approximation. If we take $|p| = 2\log 2 = 1.386$, $\rho^{-2} = 39/8$, or $\rho = 0.453$, approximately. The length of the part of the tube for which the radius of curvature lies between 0.447 and 0.453 is

$$2\int_0^{1/4} (1 - u^2)^{-1}[1 - n^2u^{2n-2}(1 - u^2)^2(1 - u^{2n})^{-2}]^{\frac{1}{2}}\,du,$$

which equals 0.511 approximately for $n \geq 5$. Since for $n = 5$, the whole length of the tube is 2.59, nearly one-fifth of this length has a radius of curvature between 0.447 and 0.453.

As n increases, the ratio of this part to the total length diminishes to zero, but for $n = 100$ it is still about 1/11. Hence it appears that local overlapping may be serious for values of $\cos\theta$ appreciably less than the critical value.

Nonlocal overlapping will not occur. It is necessary for such overlapping that the tube should bend around so that two points P_1 and P_2 of the axis are at a geodesic distance apart less than twice the radius of the tube even though they are separated by a considerably greater distance than this, measured along the axis.

If P_1 and P_2 correspond to values q_1 and q_2 of the parameter q, the square of the distance in Euclidean n-space between them is given by

$$D = \sum_\alpha (\lambda_2 q_2^\alpha - \lambda_1 q_1^\alpha)^2, \tag{6.13}$$

where

$$\lambda_i^2 = q_i^{-2}(1 - q_i^2)(1 - q_i^{2n})^{-1}, \qquad i = 1, 2.$$

If we transform to polar coordinates $\theta_1, \cdots, \theta_{n-1}$ on the unit hypersphere and let P_1 be the point $(0, 0, \cdots, 0)$ and P_2 the point $(\alpha, 0, \cdots, 0)$, then $D^2 = 2 - 2\cos\alpha$ and the geodesic distance between P_1 and P_2 is α. If, therefore, D is the distance from a fixed point q_1 to a variable point q,

$$D^2 = 2 - 2\,\frac{1 - q^n q_1^n}{1 - qq_1}\left\{\frac{(1 - q^2)(1 - q_1^2)}{(1 - q^{2n})(1 - q_1^{2n})}\right\}^{\frac{1}{2}}, \tag{6.14}$$

and a minimum value of D^2 corresponds to a maximum value of

$$\cos^2\alpha = \left(\frac{1 - q^n q_1^n}{1 - qq_1}\right)^2 \frac{(1 - q^2)(1 - q_1^2)}{(1 - q^{2n})(1 - q_1^{2n})}. \tag{6.15}$$

The ends of the axis of the tube are at a geodesic distance $\pi/2$ apart. One end of the axis is at the point where the positive x_1 coordinate axis cuts the unit hypersphere and the other end is at the point where the positive x_n axis cuts it. The axis of the tube lies wholly on that part of the surface of the hypersphere for which all coordinates are positive, and so cannot spiral around the end points or form an equatorial spiral around the sphere in the middle.

If there is nonlocal overlapping there must be at least three distinct roots of (6.15), considered as an equation in q corresponding to a given q_1 and a given α (less than twice the geodesic radius of the tube). Two of the roots will represent neighbouring points on the axis, one on each side of q_1, and the others, if they exist, correspond to points on nonlocal portions of the axis. If the point q_1 is at a geodesic distance less than α from either end of the axis of the tube, the existence of *two* distinct roots would imply nonlocal overlapping.

Since the tube is symmetrical about the middle of its axis (at $q = 1$) we can assume $0 \leq q_1 \leq 1$. Then q can take any real value from 0 to ∞. By the condition for the absence of *local* overlapping, $\cos^2 \alpha > 0.360$.

If $q_1 = 0$, the equation for q becomes

$$1 + q^2 + q^4 + \cdots + q^{2n-2} - \sec^2 \alpha = 0, \tag{6.16}$$

and this, by Descartes' rule of signs, has at most one real root. There is therefore no nonlocal overlapping at the ends of the tube, for any value of n. This is also true at the middle, where $q_1 = 1$.

For any finite n the geodesic distance β of q_1 from the end $q = 0$ is given by $1 - q_1^2 = (1 - q_1^{2n}) \cos^2 \beta$. If there is to be no nonlocal overlapping, the equation in q,

$$\left(\frac{1 - q^n q_1^n}{1 - qq_1}\right)^2 \left(\frac{1 - q^2}{1 - q^{2n}}\right) = \frac{\cos^2 \alpha}{\cos^2 \beta}, \tag{6.17}$$

should possess only one real positive root if $\beta < \alpha$ and only two real positive roots if $\beta > \alpha$. If $\beta = \alpha$, one root is $q = 0$. The equation in this case becomes

$$1 + q^2 + \cdots + q^{2n-2} = (1 + qq_1 + \cdots + q^{n-1}q_1^{n-1})^2. \tag{6.18}$$

Writing $y_1 = 1 + q^2 + \cdots + q^{2n-2}$ and $y_2 = (1 + qq_1 + \cdots + q^{n-1}q_1^{n-1})^2$, it is clear that y_1 and y_2 and all their derivatives are nonnegative for all q, that y_1 and y_2 are never zero, and that at $q = 0$, $y_2' > 0$. The curve of y_2 as a function of q starts together with that of y_1 at $q = 0$ and remains above the latter for an interval of $q > 0$. When q tends to infinity, $y_2/y_1 \to q_1^{2n-2}$, which is less than 1, so that the curves must eventually cross again. A real root of $y_1 = y_2$ greater than zero therefore exists. We shall now show that this root is unique.

Since $y_1 < y_1^2 < y_2$ for $0 < q < q_1$, we can confine our attention to the case $q > q_1$.

If $q_1 = 1$, (6.18) clearly has no real root for $q > 0$. Moreover, y_2 is a continuous function of q_1. Hence, if for any q_1, between 0 and 1, two real roots exist, there must be some q_1 and some corresponding q such that (6.18) has a *double* root.

It is sufficient therefore to show that such a double root cannot exist. The conditions are

$$y_1 - y_2 = 0, \qquad y_1' - y_2' = 0. \tag{6.19}$$

Writing $y_1' = [n - q^2(1 + q^2 + \cdots + q^{2n-2})]/(1 - q^2)$ and $y_2' = \sqrt{y_2}[n - qq_1(1 + qq_1 + \cdots + q^{n-1}q_1^{n-1})]/(1 - qq_1)$, the second condition of (6.19) gives

$$\frac{n - q^2 - q^4 - \cdots - q^{2n}}{n - qq_1 - q^2q_1^2 - \cdots - q^nq_1^n} \cdot \frac{1 - qq_1}{1 - q^2} = \sqrt{y_2}. \tag{6.20}$$

From the first condition,

$$\sqrt{y_2} = \frac{y_1}{\sqrt{y_2}} = \frac{1 + q^2 + \cdots + q^{2n-2}}{1 + qq_1 + \cdots + q^{n-1}q_1^{n-1}}.$$

Substituting for $\sqrt{y_2}$ in (6.20), subtracting 1 from both sides, removing a common factor $q^2 - qq_1$, cross-multiplying and collecting terms, we arrive at an equation of the form

$$A + Bqq_1 + Cq^2q_1^2 + \cdots + Zq^{n-2}q_1^{n-2} = 0,$$

where all the coefficients are positive. This can obviously not be satisfied for positive q and q_1.

In the more general case of (6.17) the equation is $y_1 = cy_2$, where $c = \cos^2\beta/\cos^2\alpha$, and it is readily verified that the above argument holds for $c > 1$. If $c < 1$, the curve for cy_2 is below that for y_1 at $q = 0$ and at $q = \infty$, so that if a real root exists at all there will be at least two roots. These roots cannot coincide, since if they did we should have $y_1 = cy_2$ and $y_1' = cy_2'$ simultaneously, which is ruled out by the above argument. The same argument also shows that there cannot be more than two roots. That there are at least two follows from the fact that when $y_1 = \sec^2\beta$, say at $x = b$, $cy_2 = \sec^2\beta\sec^2\alpha$, so that $y_1/cy_2 = 1/\sec^2\alpha < 1$. The curve for y_1 is therefore below that of cy_2 at $x = b$.

For the modified exponential curve the expression for $\cos^2\alpha$ corresponding to (6.15) is

$$\cos^2\alpha = \left\{\frac{1 - q^nq_1^n}{1 - qq_1} - \frac{(1 - q_1^n)(1 - q^n)}{n(1 - q_1)(1 - q)}\right\}^2 \Big/ ff_1, \tag{6.21}$$

where

$$f = \frac{1 - q^{2n}}{1 - q^2} - \frac{1}{n}\left(\frac{1 - q^n}{1 - q}\right)^2$$

and f_1 is the same expression with q_1 instead of q.

When $q_1 = 1$,

$$\cos^2\alpha = \frac{3}{n} \cdot \frac{n - 1}{n + 1}\left(\frac{1 - q^n}{1 - q}\right)^2 \Big/ f, \tag{6.22}$$

which may be written

$$\frac{1+q^n}{1+q} = \frac{1-q^n}{1-q}\frac{1}{n}\{1+3(n-1)\sec^2\alpha/(n+1)\}$$

or

$$(1+q)(1+q+q^2+\cdots+q^{n-1})\left(\frac{1}{n}+\frac{3(n-1)\sec^2\alpha}{n(n+1)}\right)-(1+q^n)=0.$$

Since in this equation there are only two changes of sign, the factor $\frac{1}{n}+\frac{3(n-1)\sec^2\alpha}{n(n+1)}$ being less than 1 for admissible values of α, there cannot be more than two real positive roots. Hence there is no nonlocal overlapping near the middle of the tube.

When $q_1 = 0$,

$$\cos^2\alpha = \frac{n}{n-1}\left\{1-\frac{1-q^n}{n(1-q)}\right\}^2\bigg/ f. \tag{6.23}$$

It may be shown that the derivative of the right-hand side, considered as a function of q, is negative for all values of $q > 0$. Since the right-hand side is equal to 1 for $q = 0$ and to $(n-1)^{-2}$ for $q = \infty$, there is just one real positive root for any admissible value of $\cos^2\alpha$. Therefore no nonlocal overlapping is possible at the ends of the tube.

Moreover, it is readily shown that $\cos^2\alpha$ in (6.22) has no maximum or minimum for any value of q except 0 or 1. That is, the geodesic distance from the midpoint of the tube to a variable point P of the axis increases monotonically as P moves away towards either end. The same conclusion follows from (6.23) as P moves away from the end of the tube towards the middle. This circumstance, which of course holds also for the simple exponential tube, suggests that the possibility of nonlocal overlapping is effectively ruled out.

7. Probability formulas and tables. The "volume" of a tube of geodesic radius θ surrounding the projected curve will be given by

$$l_n\pi^{\frac{1}{2}(n-2)}\sin^{n-2}\theta/\Gamma(\tfrac{1}{2}n), \tag{7.1}$$

where l_n is a function of n evaluated in Section 5. For any value of $n > 2$ there will also be hemispherical "caps" at the ends of the curve. The "volume" of a complete cap of radius θ surrounding a given point on the hypersphere is

$$\begin{aligned}\frac{2\pi^{\frac{1}{2}(n-2)}}{\Gamma[\frac{1}{2}(n-2)]}\int_0^{2\pi\sin(\theta/2)}\int_0^{\theta/2}\cos\varphi\sin^{n-3}\varphi\,d\varphi\,ds \\ = 2\pi^{\frac{1}{2}n}\sin^{n-1}(\theta/2)/\Gamma(\tfrac{1}{2}n),\end{aligned} \tag{7.2}$$

and we may consider this as the sum of the two hemispherical caps at the ends.

The probability that a random sample point will lie within the tube is therefore given by

$$(7.3)\qquad p(\theta) = \frac{l_n}{2\pi}\sin^{n-2}\theta + \sin^{n-1}(\theta/2).$$

In terms of the correlation coefficient R, the probability of obtaining by chance a value of R at least as great as R_0 is

$$(7.4)\qquad p(R_0) = \frac{l_n}{2\pi}(1 - R_0^2)^{\frac{1}{2}(n-2)} + \{\tfrac{1}{2}(1 - R_0)\}^{\frac{1}{2}(n-1)}.$$

This, of course, is true only for values of $R_0 > 0.894$. It will often happen, however, that when the data suggest an exponential trend the correlation between observed and fitted values will be high.

The value of R_0 corresponding to an assigned significance level can be calculated from equation (7.4) for particular values of n. A few such values are given in Table III, where it will be noted that in each column the last entry is below the critical value.

TABLE III

Values of correlation coefficient corresponding to certain significance levels ($Y = be^{pz}$)

n	Significance level .05	.01	.001	.0001
3	.990	.9995	>.9999	>.9999
4	.938	.987	.999	.9999
5	.876	.958	.991	.9996
6		.923	.976	.992
7		.887	.956	.983
8			.935	.970
9			.912	.955
10			.889	.939
12				.906
15				.859

For the modified exponential, the probability of obtaining by chance a value of the correlation coefficient R between observed and computed values of y at least as great as R_0, on the null hypothesis that $b = 0$ in the parent population, is given by

$$(7.5)\qquad p(R_0) = \frac{l_n'}{2\pi}(1 - R_0^2)^{\frac{1}{2}(n-3)} + \{\tfrac{1}{2}(1 - R_0)\}^{\frac{1}{2}(n-2)},$$

for $R_0 > 0.894$. This is the same formula as (7.4) with $n - 1$ instead of n, because of the loss of one dimension in projection. Also, l_n' is now given by (5.12), instead

of (5.7). A short table, computed on the basis of this formula and assuming that no overlapping exists, is appended as Table IV.

TABLE IV

Correlation coefficient corresponding to certain significance levels ($Y = a + be^{px}$)

n	Significance level .05	.01	.001	.0001
4	.983	.999	.9999	.9999
5	.918	.983	.998	.9998
6	.849	.949	.989	.998
7		.910	.972	.991
8		.872	.951	.981
9			.904	.952
10			.881	.935
11				.918
12				.901

Note that for Table III the correlation coefficient is computed *without* elimination of the means, whereas for Table IV the correlation coefficient is computed in the usual way.

8. Methods of fitting curves. As n increases, smaller values of R become significant at any given level, and for moderately large n these values of R, at the lower levels, pass out of the region for which the calculation is valid. However, the table may be useful in deciding whether an assumed exponential regression is plausible, when the number of sample points available is fairly small.

Tables for use in fitting exponential curves may be found in Glover's *Tables* [3] but unfortunately these tables cover a very limited range of values of p (from 0 to 0.0953, e^p between 1.0 and 1.1). The exact solution by least squares is laborious. Values of b and p are calculated from the equations

$$\begin{aligned} \sum_{\alpha} ye^{p\alpha} &= b\Sigma e^{2p\alpha}, \\ \Sigma\alpha ye^{p\alpha} &= b\Sigma\alpha e^{2p\alpha}. \end{aligned} \tag{8.1}$$

Rough approximations to b and p may be found by fitting a straight line to the values of $\log y$, and these approximations may be improved by Seidel's method.

Villars [4] has recently given some approximate methods of fitting the modified exponential equation $y = a + be^{px}$. In the first method the n observations (n even) are divided into two groups, one including the 1st, 3rd, 5th, etc., and the other the 2nd, 4th, 6th, etc. The relationship between the expectations of corresponding members of the two groups is

$$\nu_j = h + m\mu_j, \qquad j = 1, 2, \cdots, n/2, \tag{8.2}$$

where

$$\mu_j = a + be^{(2j-1)p}, \tag{8.3}$$
$$\nu_j = a + be^{2jp},$$

and so $m = e^p$, $h = a(1 - e^p)$. Hence if a straight line through the origin is fitted to the observed u and v values, where $u_j = y_{2j-1}$, $v_j = y_{2j}$, both variables being subject to error, the slope of this line will give an estimate of e^p and its intercept will give an estimate of $a(1 - e^p)$. An estimate of b can then be found from (8.5), or alternatively both a and b can be found from (8.4) and (8.5), where

$$Na + b\,\Sigma m^\alpha = \Sigma y, \tag{8.4}$$

$$a\,\Sigma m^\alpha + b\,\Sigma m^{2\alpha} = \Sigma y m^\alpha. \tag{8.5}$$

An alternative method, also given by Villars, is applicable whether n is odd or even. It consists in treating y_j and y_{j+1}, $j = 1, 2, \cdots, n - 1$, as pairs of corresponding u and v values, although, since each u except the last appears also as a v, the pairs are clearly not independent.

A systematic method of calculating the exact least squares solution, starting with an approximate value of p (or of $q = e^p$), has been presented by W. J. Spillman [7]. This method utilises tables of q^x for x between 2 and 20 and for q at intervals of .01 between 0 and 1.

There is, of course, no point in fitting an exponential equation by least squares, or by any more or less equivalent method, unless there is some reason to believe that the underlying assumption of approximate uniformity in the variance of y is valid. If $\log y$ is approximately normal with constant standard deviation, as seems to be true for many data in the field of economics, the usual procedure of fitting a straight line to the logarithms of the values of y is clearly justified. On the other hand, if the standard deviation of y is constant, the effect of this procedure is to give undue weight to the smaller values of y. Some data exist on fertilizer trials in which the assumption of constant standard deviation of y seems reasonable, and which suggest an exponential, or modified exponential, trend.

9. Numerical illustrations. The data in Table V, referring to the mean girths y of rubber trees in inches after various levels x of fertilizer treatment, I owe to the courtesy of Mr. H. Fairfield Smith.

TABLE V

x	y	Y
0	20.518	20.526
1	21.138	21.109
3	21.734	21.804
5	22.218	22.144
7	22.286	22.311

It will be observed that the values of x are not equidistant. This makes the calculations more awkward.

If the fitted equation is $Y = a + bq^x$, where $q = e^p$, the least squares equations for a, b, q are

$$\begin{aligned} 5a + b[1 + q + q^3 + q^5 + q^7] &= y_0 + y_1 + y_3 + y_5 + y_7\,, \\ a[1 + q + q^3 + q^5 + q^7] + b[1 + q^2 + q^6 + q^{10} + q^{14}] & \\ &= y_0 + y_1 q + y_3 q^3 + y_5 q^5 + y_7 q^7, \\ a[1 + 3q^2 + 5q^4 + 7q^6] + b[q + 3q^5 + 5q^9 + 7q^{13}] & \\ &= y_1 + 3y_3 q^2 + 5y_5 q^4 + 7y_7 q^6. \end{aligned} \tag{9.1}$$

Approximate values for a, b, and q were found by Cowden's method [6]. A curve was drawn by eye between the plotted points and a trial value of a calculated from three equidistant ordinates, Y_0, Y_1, Y_2, by the formula

$$a = \frac{Y_0 Y_2 - Y_1^2}{Y_0 + Y_2 - 2Y_1}\,. \tag{9.2}$$

Values of $Y - a$ were then plotted on semi-logarithmic paper, and the value of a was adjusted by trial so that a straight line fitted the points reasonably well. From this straight line, values of b and q were obtained, b being the ordinate at $x = 0$, and q^7 the ratio of the ordinates at $x = 7$ and $x = 0$. In this way the following approximate values were calculated:

$$a_0 = 22.5, \qquad b_0 = -2.0, \qquad q_0 = 0.70.$$

One application of Seidel's method (solving linear equations in δa, δb, δc) gave the improved values

$$a = 22.47, \qquad b = -1.945, \qquad q = 0.7000.$$

Using these values, the calculated Y of Table V were obtained. The correlation coefficient between observed y and computed Y is 0.99735, which corresponds to $n = 5$. If, ignoring the slight deviation from uniformity in the x-intervals, we use Table IV, we find that P is slightly >0.001. The null hypothesis of no effect of the fertilizer is decisively rejected.

Villars [4] gave an illustration of the fitting of an exponential curve to data referring to a certain property of a rubber latex. By his first method he obtained as the regression equation

$$Y = 1.0009 - 0.2877e^{-0.2236x}, \tag{9.3}$$

where $2x = t + 1$ of his formula (4.1). By the second method he obtained the equivalent of

$$Y = 1.0000 - 0.2811e^{-0.2218x}. \tag{9.4}$$

If the correlation coefficients between Y and y are computed for both equations, the values turn out to be 0.9560 and 0.9572, respectively, so that the second

method gives a slightly better fit. The number of observations, however, is large enough (sixteen) for such a coefficient of correlation to be very highly significant, with reference to the null hypothesis.

For the purposes of illustration, we will use only the first six of Villars' observations, given in the first two columns of Table VI.

TABLE VI

x	y	Y (method 1)	Y (method 2)
1	0.776	0.7715	0.7742
2	0.852	0.8555	0.8415
3	0.850	0.8826	0.8754
4	0.869	0.8914	0.8924
5	0.939	0.8942	0.9010
6	0.904	0.8951	0.9053

By Villars' first method the values of Y given in column 3 were calculated, corresponding to the equation $Y = 0.8955 - 0.3851e^{-1.1329x}$. The correlation coefficient between y and Y is $R = 0.871$, so that from Table IV the departure from the null hypothesis is barely significant. Villars' second method gives the values in column 4, corresponding to $Y = 0.9088 - 0.2693e^{-0.6870x}$, and in this case $R = 0.906$. The fit is therefore appreciably better, and the regression appears to be significant at a level about midway between .01 and .05.

10. Acknowledgement. In conclusion, the author would like to express his great indebtedness to Professor Harold Hotelling for having suggested the problem and for helpful advice and criticism.

REFERENCES

[1] Harold Hotelling, "Tubes and spheres in n-spaces and a class of statistical problems," *Am. Jour. of Math.*, Vol. 61 (1939), p. 440.

[2] D. M. Starkey, "The distribution of the multiple correlation coefficient in periodogram analysis," *Annals of Math. Stat.*, Vol. 10 (1939), p. 327.

[3] J. W. Glover, *Tables of Applied Mathematics*, George Wahr, Ann Arbor, Mich., 1923, pp. 468–481.

[4] D. S. Villars, "A significance test and estimation in the case of exponential regression," *Annals of Math. Stat.*, Vol. 18 (1947), p. 596.

[5] D. S. Villars and T. W. Anderson, "Significance tests for normal bivariate distributions," *Annals of Math. Stat.*, Vol. 14 (1943), p. 141.

[6] D. J. Cowden, "Simplified methods of fitting certain types of growth curves," *Jour. Am. Stat. Assn.*, Vol. 42 (1947), p. 585.

[7] W. J. Spillman, *Use of the Exponential Yield Curve in Fertilizer Experiments* (Technical Bulletin No. 348), U. S. Department of Agriculture, Washington, D. C.

ON THE DURATION OF RANDOM WALKS[1]

By Wolfgang Wasow

Institute for Numerical Analysis

Summary and introduction. In a recent paper [1] the author investigated the mean number of steps in random walks in n-dimensional domains. The purpose of the present article is to generalize those results by applying similar methods to the study of the moment generating function for the number of steps and of its distribution function. As an application explicit asymptotic expressions for the variance in special cases and estimates for the likelihood of very long walks are obtained.

The author wishes to express his thanks to Professor R. Fortet for many helpful discussions.

The walks take place in an open bounded domain B of n-dimensional Euclidean space E with boundary C. A point moves in E according to a given transition probability law $F(y, x)$. Here x and y are points of E with coordinates x_i, $i = 1, 2, \cdots, n$, and y_i, $i = 1, 2, \cdots, n$, and $F(y, x)$ is the probability that a jump known to start at x end at a point all of whose coordinates are less than the corresponding ones of y. The function $F(y, x)$ is a distribution function with respect to y, and it is assumed to be Borel measurable with respect to all variables. Let $N = N_x$ be the number of steps in a random walk that begins at a point x of B and ends with the step on which the moving point leaves B for the first time. If the probability of the moving point eventually leaving B is equal to one, then N is a random variable. It is called the *duration* of the walk. It is useful to extend the definition of N by setting

$$N_x = 0, \qquad x \varepsilon E - B.$$

1. The moment generating function of the duration. The probability distribution of the duration, given by the functions

$$p_k(x) = Pr\{N_x = k\}, \qquad k = 0, 1, 2, \cdots, \tag{1.1}$$

satisfies the recursion relations

$$\begin{aligned} p_{k+1}(x) &= \begin{cases} \displaystyle\int_E p_k(y)\, dF(y, x), & x \varepsilon B, \\ 0, & x \varepsilon E - B, \end{cases} \\ p_0(x) &= \begin{cases} 0, & x \varepsilon B, \\ 1, & x \varepsilon E - B. \end{cases} \end{aligned} \tag{1.2}$$

Here and in the sequel all Stieltjes differentials are formed with respect to the first argument.

---

[1] The preparation of this paper was sponsored in part by the Office of Naval Research.

We need some hypothesis sufficient to ensure the relation $\sum_{k=0}^{\infty} p_k(x) = 1$ and the existence of the moment generating function of N_x. This is the purpose of

ASSUMPTION A. *There exists a positive integer m and a positive number c < 1, both independent of x, such that*

$$Pr\{N_x \geq m\} \leq c$$

for all x in B.

From this assumption, which is slightly more general than the corresponding condition in [1], the equality $\sum_{k=0}^{\infty} p_k(x) = 1$ follows by a simple argument similar to that in [2], pp. 431–432. In fact, the last inequality implies

$$Pr\{N_x \geq jm\} \leq c^j$$

and therefore

$$\lim_{j\to\infty} \sum_{k=0}^{mj} p_k(x) = \lim_{j\to\infty} [1 - Pr\{N_x \geq jm\}] \geq \lim_{j\to\infty} (1 - c^j) = 1.$$

The aim of this section is the following theorem:

THEOREM 1. *If Assumption A is satisfied, then the moment generating function* $\phi(s, x) = \sum_{j=0}^{\infty} e^{sj} p_j(x)$ *of the duration* N_x *exists in a complex neighborhood of* $s = 0$ *and is the unique solution of the integral equation problem*

$$\phi(s, x) = \begin{cases} e^s \int_B \phi(s, y)\, dF(y, x), & x \in B, \\ 1, & x \in E - B. \end{cases} \tag{1.3}$$

Because of later need we prove a slightly more general result.

LEMMA 1. *If Assumption A is satisfied and f(x) is a real function such that* $|f(x)| \leq K$ *in* $E - B$, *then the integral equation problem*

$$u(s, x) = \begin{cases} e^s \int_B u(s, y)\, dF(y, x), & x \in B, \\ f(x), & x \in E - B, \end{cases} \tag{1.4}$$

possesses for $Re\ s < s_0 (s_0 > 0)$ *a unique solution. This solution satisfies the inequality*

$$|u(s, x)| \leq \phi(Re\ s, x)\cdot K, \qquad x \in B, \tag{1.5}$$

where $\phi(s, x)$, *the moment generating function of the duration, is the solution of* (1.3).

PROOF. Assume, at first, that $f(x)$ is nonnegative, and that s is real. Set

$$u_0(x) = \begin{cases} 0, & x \in B, \\ f(x), & x \in E - B, \end{cases}$$

$$u_{k+1}(x) = \begin{cases} e^s \int_B u_k(y)\, dF(y, x), & x \in B, \\ f(x), & x \in E - B. \end{cases} \tag{1.6}$$

Then

$$(1.7)\qquad u_k(x) = \sum_{j=1}^{k} e^{js} q_j(x), \qquad x \in B, \qquad k = 1, 2, \cdots,$$

where, for x in B,

$$(1.8)\qquad q_1(x) = \int_{E-B} f(y)\, dF(y, x),\ q_{j+1}(x) = \int_B q_j(y)\, dF(y, x).$$

The $q_j(x)$ are nonnegative, and the $u_k(x)$ form therefore a nondecreasing sequence. Iterating (1.6) m times we find, for $k > m$,

$$(1.9)\qquad u_{k+1}(x) = e^{ms} \int_B u_{k-m}(x)\, dF_m(y, x) + \chi_m(x), \qquad x \in B,$$

where

$$F_1(y, x) = F(y, x), \qquad F_m(y, x) = \int_B F_{m-1}(y, z)\, dF(z, x),$$

and the $\chi_m(x)$ are bounded and nonnegative. If x is in B, then, by Assumption A,

$$\int_B dF_m(y, x) = Pr\{N_s > m\} \le c < 1.$$

Let

$$L_k = \underset{x \in B}{\text{l. u. b.}}\ u_k(x), \qquad M = \underset{x \in B}{\text{l. u. b.}}\ \chi_m(x).$$

(L_k and M depend on s.) Then, by (1.9),

$$L_{k+1} \le e^{ms} L_k c + M.$$

Hence,

$$L_{k+1} \le M/(1 - e^{ms}c),$$

and, therefore, the nondecreasing sequence $u_k(x)$ is bounded for all s for which $e^{ms}c < 1$. Thus, it tends to a limit $u(x) = u(s, x)$, which satisfies (1.4) and can be written, by (1.7), in the form

$$(1.10)\qquad u(s, x) = \sum_{j=1}^{\infty} e^{js}\, q_j(x).$$

Since this is a power series in e^s it converges for complex s also, as long as $Re\ s < s_0 = -\log c/m$. Furthermore, we see by comparison of (1.8) and (1.2) that for $f(x) \equiv 1$ we have $q_k(x) = p_k(x)$ and $u(s, x) = \phi(s, x)$.

To prove the uniqueness, it suffices to show that $f(x) \equiv 0$ implies $u(x) \equiv 0$. By iteration of (1.4) we find, for $f(x) \equiv 0$,

$$u(x) = e^{ms} \int_B u(y)\, dF_m(y, x),$$

and hence

$$\operatorname*{l.\,u.\,b.}_{x \varepsilon B} | u(x) | \leq | e^{ms} | c \cdot \operatorname*{l.\,u.\,b.}_{y \varepsilon B} | u(y) | .$$

For values of s such that $| e^{ms} | c < 1$ this implies, indeed, that $u(x) \equiv 0$.

If $f(x)$ is not necessarily positive, then we consider the integral equation problems

$$u^{(1)} = \begin{cases} e^s \int_B u^{(1)} \, dF & \text{in } B, \\ K & \text{in } E - B, \end{cases}$$

$$u^{(2)} = \begin{cases} e^s \int_B u^{(2)} \, dF & \text{in } B, \\ K - f(x) & \text{in } E - B, \end{cases}$$

which do have unique solutions by what has been proved already. $u = u^{(1)} - u^{(2)}$ is therefore the unique solution of the original problem. We note that this argument also extends the validity of the formulas (1.8) and (1.10) to the case that $f(x)$ is not necessarily positive.

Finally, the inequality (1.5) follows easily from (1.8), (1.2), and (1.10), since

$$| u(x) | \leqq \sum_{j=1}^{\infty} e^{jRe\, s} | q_j(x) | \leqq K \sum_{j=1}^{\infty} e^{jRe\, s} p_j(x) = K\phi(Re\, s, x).$$

This completes the proof of Lemma 1. Theorem 1 implies that all moments $M_k(x)$ of N_x exist.

THEOREM 2. *The* kth *moment $M_k(x)$ of the duration N_x satisfies, for $k > 0$, the integral equation problem*

$$(1.11) \qquad M_k(x) = \begin{cases} \int_B M_k(y) \, dF(y, x) + f_k(x), & x \varepsilon B, \\ 0, & x \varepsilon E - B, \end{cases}$$

where

$$(1.12) \qquad f_k(x) = \sum_{j=1}^{k} (-1)^{j-1} \binom{k}{j} M_{k-j}(x).$$

PROOF. From the definition

$$M_k(x) = \sum_{j=0}^{\infty} j^k p_j(x)$$

follows by an application of (1.2) that, for x in B,

$$(1.13) \qquad \int_B M_k(y) \, dF(y, x) = \sum_{j=1}^{\infty} (j - 1)^k p_j(x).$$

Expansion of the binomial expression in the second member followed by an interchange of summations proves the theorem.

These integral equations for the moments form an inductive sequence, since $f_k(x)$ depends only on the moments of order $j < k$. The equation for $M_1(x)$ was the main subject of [1].

2. An asymptotic differential equation for the moment generating function. In the important case that the transition probability $F(y, x)$ is strongly concentrated about x, the integral equation (1.3) will now be shown to be approximately equivalent to a differential equation. To do this it will be assumed, as in [1] and [2], that $F(y, x) = F(y, x, \mu)$ depends on a small positive parameter μ in such a way that the following three hypotheses are satisfied.

ASSUMPTION B. *Denote by $a_i(x, \mu)$, $b_{ik}(x, \mu)$, $i, k = 1, 2, \cdots, n$, the first and second moments of $F(y, x, \mu)$ about x. Then*

$$a_i(x, \mu) = \alpha_i(x)\mu + o(\mu), \tag{2.1}$$

$$b_{ik}(x, \mu) = \beta_{ik}(x)\mu + o(\mu). \tag{2.2}$$

These relations hold uniformly for x in B. The $\alpha_i(x)$ and $\beta_{ik}(x)$ are twice continuously differentiable in $B + C$.

ASSUMPTION L. *Let $K_r(x)$ denote the sphere of radius r with center at x. Then*

$$\int_{E-K_r(x)} (y_i - x_i)(y_k - x_k)\, dF(y, x, \mu) = o(\mu), \qquad i, k = 1, 2, \cdots, n,$$

uniformly with respect to x in B, for any fixed $r > 0$.

These assumptions could very likely be weakened to the equivalent of the analogous hypotheses in [12].

ASSUMPTION E. *The matrix $\{\beta_{ik}(x)\}$, which is obviously nonnegative, is positive definite in $B + C$.*

For what follows we also require a certain degree of smoothness of the boundary.

ASSUMPTION S. *The boundary C has a continuously turning tangent plane.* (This restriction could be considerably weakened; e.g., by inserting the word "piecewise", cf. [2], p. 438.)

We prove first

LEMMA 2. *Assumptions B, L, and E imply Assumption A, at least for sufficiently small μ.*

PROOF. To simplify the notation we give the proof first for the one-dimensional case, in which we can write α, β, x, y for $\alpha_i, \beta_{ii}, x_i, y_i$. Using Assumptions B and L we have, for any $\epsilon > 0$,

$$\beta\mu = \int_{-\infty}^{\infty} (y - x)^2\, dF(y, x, \mu) + o(\mu) = \int_x^{x+\epsilon} (y - x)^2\, dF$$

$$+ \int_{x-\epsilon}^{x} (y - x)^2\, dF + o(\mu) \leq \epsilon \int_x^{x+\epsilon} (y - x)\, dF - \epsilon \int_{x-\epsilon}^{x} (y - x)\, dF + o(\mu)$$

and also

$$\alpha\mu = \int_x^{x+\epsilon} (y - x)\, dF + \int_{x-\epsilon}^{x} (y - x)\, dF + o(\mu).$$

Multiplying the last equality by ϵ and adding it to the preceding inequality, we find

$$(\beta - \epsilon\alpha)\mu + o(\mu) \leq 2\epsilon \int_x^{x+\epsilon} (y - x)\, dF.$$

Since $\beta(x) \geq$ const. > 0 in B, by assumption E, we see, by first choosing ϵ sufficiently small, that for small μ

$$\int_x^{x+\epsilon} (y - x)\, dF \geq C \cdot \mu, \qquad x \,\varepsilon\, B,$$

where C is an (arbitrary) positive constant.

On the other hand, for $\mu^2 < \epsilon$,

$$\int_x^{x+\epsilon} (y - x)\, dF = \int_x^{x+\mu^2} (y - x)\, dF + \int_{x+\mu^2}^{x+\epsilon} (y - x)\, dF \leq \mu^2 + \epsilon \int_{x+\mu^2}^{\infty} dF.$$

Combining this with the preceding inequality, we have

$$Pr\{y - x \geq \mu^2\} \geq C_1\mu, \qquad x \,\varepsilon\, B,$$

where C_1 is another constant. Hence Assumption A is satisfied, if m is chosen greater than the diameter D of B divided by μ^2, and $c = 1 - (C_1\mu)^{D/\mu^2}$.

In more than one dimension the same argument can be applied in any one of the coordinate directions. The changes necessary concern only the notation. Thus the lemma is proved.

Denote by $\psi(s, x, \mu)$ the moment generating function of the random variable

$$t = t_x = \mu N_x .$$

Obviously

$$\psi(s, x, \mu) = \phi(\mu s, x, \mu). \tag{2.3}$$

Let furthermore $L[u]$ be an abbreviation for the operator

$$L[u] = \tfrac{1}{2} \sum_{i,k=1}^{n} \beta_{ik}(x) \frac{\partial^2 u}{\partial x_i \partial x_k} + \sum_{j=1}^{n} \alpha_j(x) \frac{\partial u}{\partial x_j}. \tag{2.4}$$

Then the following theorem will be proved.

THEOREM 3. *If Assumptions B, L, E, and S are satisfied, then the moment generating function $\psi(s, x, \mu)$ of the random variable $t = \mu N$ satisfies the limit relation*

$$\lim_{\mu \to 0} \psi(s, x, \mu) = u(s, x) \tag{2.5}$$

uniformly for x in B, $|s| \leq s_1$, where $u(s, x)$ is the solution of the problem

$$L[u] + su = 0, \quad x \varepsilon B, \qquad u = 1, \quad x \varepsilon E - B. \tag{2.6}$$

Before we can prove the limit relation (2.5) we have to prove separately the weaker statement that $\psi(s, x, \mu)$ remains bounded as $\mu \to 0$.

LEMMA 3. *If Assumptions B, L, E, and S are satisfied, then there exist two positive constants C and C', independent of μ, such that*

$$|\psi(s, x, \mu)| \leq C \quad \textit{for} \quad |s| \leq C'. \tag{2.7}$$

PROOF. From the results of [1], in particular from Lemma 3, Theorem 1, and Theorem 2 of that paper, it follows that the solution of the problem

$$M(x) = \begin{cases} \int_B M(y)\, dF(y, x, \mu) + f(x), & x \varepsilon B, \\ 0, & x \varepsilon E - B, \end{cases} \tag{2.8}$$

satisfies the inequality

$$|M(x)| \leq C_1 \operatorname*{l.u.b}_{x \varepsilon B} |f(x)|/\mu, \tag{2.9}$$

where C_1 is a constant independent of μ. This remains true if $f(x)$ depends boundedly on μ. We wish to apply this inequality to the integral equations (1.11). To that end we first note that

$$|f_k(x)| \leq kM_{k-1}(x), \tag{2.10}$$

for $f_k(x)$ can be written, by (1.11) and (1.13), in the form

$$f_k(x) = M_k(x) - \sum_{j=1}^{\infty} (j-1)^k p_j(x) = \sum_{j=1}^{\infty} [j^k - (j-1)^k] p_j(x)$$

$$= k \sum_{j=1}^{\infty} j^{*k-1} p_j(x),$$

where $j - 1 < j^* < j$, and, therefore,

$$0 \leq f_k(x) \leq k \sum_{j=1}^{\infty} j^{k-1} p_j(x) = kM_{k-1}(x).$$

Applying (2.9) and (2.10) inductively to (1.11) we obtain

$$|M_k(x)| \leq k!(C_1/\mu)^k. \tag{2.11}$$

Substitution of this inequality into the formula

$$\psi(s, x, \mu) = \sum_{k=0}^{\infty} \frac{M_k}{k!} (\mu s)^k$$

yields

$$|\psi(s, x, \mu)| \leq \sum_{k=0}^{\infty} (C_1 s)^k = \frac{1}{1 - C_1 s}$$

for $|s| < C_1$. This proves Lemma 3.

PROOF OF THEOREM 3. The basic idea of the proof is similar to that of Theorem 2 in [1] and thus to Petrowsky's reasoning in [2]; we show that $u(x)$ satisfies an integral equation little different from that for $\psi(s, x, \mu)$, and conclude from that fact that the two functions are nearly the same.

We first replace $u(x)$ by a slightly different function $u_\delta(x)$ defined in a larger domain B'', in order to avoid extraneous difficulties near the boundary C. This can be done by constructing a twice continuously differentiable mapping

$$x_i' = f(x, \delta), \qquad x \varepsilon B, \qquad i = 1, 2, \cdots, n,$$

which is continuous in δ, for $\delta \geq 0$, together with its first and second derivatives with respect to x, and has the following properties.

(a) It reduces to the identity for $\delta \to 0$.

(b) It is, for all δ, the identity transformation in a subdomain B' of B that tends to B as $\delta \to 0$.

(c) It maps B onto a domain B'' containing B in its interior for $\delta > 0$.

For the explicit construction of such a mapping with the help of Assumption S we refer to [2].

If we define $u_\delta(x)$ in B by

$$u_\delta(x') = u(x), \qquad x \varepsilon B,$$

then this new function is defined and twice continuously differentiable in B''. It tends to $u(x)$, uniformly in B, together with its first and second derivatives. We extend the definition into the whole space E by setting

$$u_\delta(x) = 1, \qquad x \varepsilon E - B.$$

Next, it can be shown that, for any $\epsilon > 0$, we can choose first a $\delta > 0$ and then a $\mu_0 > 0$ such that, for $|s| \leq s_0$,

$$\int_E u_\delta(y)\, dF(y, x) = (1 - \mu s)u_\delta(x) + \mu g(s, x, \mu), \qquad x \varepsilon B, \tag{2.12}$$

where

$$|g(s, x, \mu)| < \epsilon, \tag{2.13}$$

provided $\mu \leq \mu_0$. The proof of this statement resembles so much the analogous arguments in [1] and [2] that it will be omitted here. (Formula (2.12) is essentially the result of expanding $u_\delta(x)$ about x by Taylor's formula up to quadratic terms and applying Assumptions B and L.)

Because of the definition of $u_s(x)$ it can also be assumed that δ has been chosen so small that

$$|u_s - u| \le \epsilon, \qquad x \,\varepsilon\, B, \tag{2.14}$$

$$|u_s - 1| \le \epsilon, \qquad x \,\varepsilon\, E - B. \tag{2.15}$$

For μ_0 sufficiently small,

$$|(1 - \mu s - e^{-\mu s})u_s| \le \mu\epsilon, \qquad x \,\varepsilon\, B, \qquad \mu \le \mu_0 .$$

Therefore we can write, instead of (2.12),

$$e^{\mu s} \int_E u_s(y)\, dF(y, x) = u_s(x) + \mu h(s, x, \mu), \qquad x \,\varepsilon\, B,$$

with

$$|h(s, x, \mu)| \le 3\epsilon. \tag{2.16}$$

We now split u_s into the sum $u_s = u_s^{(1)} + u_s^{(2)}$, where

$$u_s^{(1)} = e^{\mu s} \int_E u_s^{(1)}\, dF \quad \text{in} \quad B, \qquad u_s^{(1)} = u_s \quad \text{in} \quad E - B, \tag{2.17}$$

$$u_s^{(2)} = e^{\mu s} \int_E u_s^{(2)}\, dF + \mu h \quad \text{in} \quad B, \qquad u_s^{(2)} = 0 \quad \text{in} \quad E - B. \tag{2.18}$$

To estimate $u_s^{(1)}$ we subtract it from (1.3) with $s\mu$ substituted for s, and use (2.15), (2.3), Lemma 1, and (2.7). This yields

$$|u_s^{(1)} - \psi| \le CK\epsilon, \qquad x \,\varepsilon\, B. \tag{2.19}$$

This implies, in particular, that $u_s^{(2)}$ is bounded as $\mu \to 0$. Therefore (2.18) can be written, for sufficiently small μ_0, in the form

$$u_s^{(2)} = \int_E u_s^{(2)}\, dF + \mu(su_s^{(2)} - h^*) \quad \text{in} \quad B, \qquad u_s^{(2)} = 0 \quad \text{in} \quad E - B,$$

where

$$|h^*| = |h^*(s, x, \mu)| \le 4\epsilon.$$

Application of (2.9) yields

$$\text{l.u.b.}\, |u_s^{(2)}| \le C_1(|s|\, \text{l.u.b.}\, |u_s^{(2)}| + 4\epsilon),$$

i.e.,

$$|u_s^{(2)}| \le 4\epsilon C_1/(1 - C_1 |s|) \tag{2.20}$$

for $|s| \le s_0 < 1/C_1$. Combining (2.19) and (2.20) we obtain the inequality

$$|u - \psi| \le \text{const. } \epsilon, \qquad x \,\varepsilon\, B,$$

which proves Theorem 3.

THEOREM 4. *If Assumptions B, L, E, and S are satisfied, then the* k*th moment* $M_k(x, \mu)$ *of the duration* N_x *satisfies, uniformly in B, the relation*

$$\lim_{\mu \to 0} \mu^k M_k(x, \mu) = m_k(x), \qquad k \geq 1,$$

where the $m_k(x)$ *are defined recursively by the conditions*

$$m_0(x) = 1,$$

and, for $k \geq 1$,

$$L[m_k] + km_{k-1} = 0, \qquad x \varepsilon B,$$

$$m_k = 0, \qquad x \text{ on } C.$$

PROOF. The solution $u(x, s)$ of (2.6) is connected with the functions $m_j(x)$ by the relation

$$u(s, x) = \sum_{j=0}^{\infty} \frac{m_j(x)}{j!} s^j,$$

as can be seen by replacing $u(x, s)$ in (2.6) by its series in powers of s and collect ing coefficients of like powers. By Theorem 3 the function

$$\psi(s, x, \mu) = \sum_{j=0}^{\infty} \frac{\mu^j M_j(x)}{j!} s^j$$

tends uniformly to $u(x, s)$ as $\mu \to 0$; hence the coefficients of the first power series are the uniform limits of those of the second. This proves Theorem 4.

For $j = 1$, Theorem 4 was proved in [1].

3. An asymptotic differential equation for the distribution function of the duration. Let

$$P(t, x, \mu) = Pr\{\mu N_x \leq t\}$$

be the distribution function of the random variable μN. From Theorem 3 we conclude by means of the continuity theorem for moment generating functions (see [3]) that there exists a distribution function $Q(t, x)$ of t such that

$$\lim_{\mu \to 0} P(t, x, \mu) = Q(t, x) \tag{3.1}$$

at all continuity points of $Q(t, x)$ with respect to t, and that $u(s, x)$ is the moment generating function of $Q(t, x)$.

The probability

$$P(t, x, \mu) = \sum_{k \leq t/\mu} p_k(x) \tag{3.2}$$

satisfies, because of (1.2), the recursive relations

$$P(t + \mu, x, \mu) = \int_E P(t, y, \mu)\, dF(y, x, \mu), \qquad x \varepsilon B, \qquad t = 0, \mu, 2\mu, \cdots,$$

$$(3.3) \qquad P(0, x, \mu) = 0, \qquad x \varepsilon B,$$

$$P(t, x, \mu) = 1, \qquad x \varepsilon E - B, \qquad t > 0.$$

From these and (3.1) it is easy to obtain, *in a purely formal way*, the differential equation (3.5) for $Q(t, x)$. The same result can be made plausible by setting

$$u(s, x) = \int_0^\infty e^{st}\, dQ(t, x) \tag{3.4}$$

in (2.6) and operating *formally* on the Stieltjes differential. Our aim in this section is to give a proof of (3.5). In spite of the plausibility of the result the proof is somewhat long, because the problem combines the features of what Khintchine calls, in [4], the first and second diffusion problems.

A feasible approach to our problem, different from that of this paper, could be based on the remark that $u(s, x)$ and therefore $Q(t, x)$ depend only on the functions $\alpha_i(x)$, $\beta_{ik}(x)$, so that $F(y, x, \mu)$ can be chosen in many different ways as far as the properties of $Q(t, x)$ are concerned. (This is an instance of the "invariance principle", used systematically, in a different context, by Kac and Erdös, cf. [5].) The most natural choice for $F(y, x, \mu)$ is the one obtained from the continuous Markov process associated with the differential equation (3.5) by considering that process at discrete time values $t = 0, \mu, 2\mu, \cdots$, only. This approach has not been chosen here, first, because we wish to preserve uniformity of method, and secondly, because the theory of such Markov processes does not seem to have been established in sufficient completeness for the n-dimensional case. (In one dimension, the proof that such a continuous process exists was given by Feller in [12]. This has been partially generalized to n dimensions by Dressel [6]. The proof that the duration of this continuous process satisfies the differential equation of (3.5) in the one-dimensional case is contained in the article [13] by Fortet.)

THEOREM 5. *If the assumptions B, L, E, and S are satisfied, then*

$$\lim_{\mu \to 0} Pr\{\mu N_x \le t\} = Q(t, x)$$

exists and is the solution of the differential equation problem

$$\begin{aligned} L[Q] - \frac{\partial Q}{\partial t} &= 0, \quad t > 0, \quad x \in B, \\ Q &= 1, \quad t > 0, \quad x \text{ on } C, \\ Q &= 0, \quad t = 0, \quad x \in B. \end{aligned} \tag{3.5}$$

The convergence is uniform in $B + C$, for any interval $0 < t_0 \le t \le t_1$.

The proof of this theorem is based on two lemmas.

LEMMA 4. *The distribution function $Q(t, x)$ is a continuous function of all arguments combined, for $t > 0$, $x \in B + C$, and for $t = 0$, $x \in B$.*

PROOF. Let

$$g(t, x, T) = \frac{1}{2T} \int_{-T}^{T} u(is, x) e^{-ist}\, ds. \tag{3.6}$$

It is known that this function is real and tends to a limit as $T \to \infty$. For fixed x the distribution $Q(t, x)$ is continuous in t if and only if

$$\lim_{T \to \infty} g(t, x, T) = 0. \tag{3.7}$$

(See, e.g., [7], p. 24, for these statements.) Also,

$$| g(t, x, T)| \leq 1, \tag{3.8}$$

since $u(is, x)$, as a characteristic function, is numerically less than one. By (2.6) the function $g(t, x, T)$ satisfies for all T the differential equation

$$L[g] - \frac{\partial g}{\partial t} = 0, \qquad x \varepsilon B, \qquad -\infty < t < \infty. \tag{3.9}$$

From (3.8) and (3.9) we can conclude (cf. [8], p. 383–384) that $\partial g/\partial t$ is uniformly bounded for all T in any finite t-interval and for x in any closed subdomain of B. Therefore the limit of g, as $T \to \infty$, is a continuous function of t. On the other hand, since $Q(t, x)$ as a distribution function has at most a denumerable set of discontinuities, $\lim_{T \to \infty} g$ is zero for fixed x, except possibly at a denumerable set of t-values. Being continuous, the limit of g must therefore be zero *everywhere* in the domain considered, i.e., $Q(t, x)$ is for all x in B and for all t a continuous function of t. (The result of Gevrey [8], referred to above, is proved in that paper only for differential equations whose second order terms form Laplace's operator. A generalization sufficient for our needs can be established by combining Gevrey's arguments with the results of [6].) In $E - B$ the distribution functions $P(t, x, \mu)$—and therefore $Q(t, x)$, their limit as $\mu \to 0$— are identically 1 for $t > 0$. Hence $Q(t, x)$ is, for $t > 0$, a continuous function of t in the closed domain $B + C$.

To prove that $Q(t, x)$ is continuous in x also, it suffices to remember that $u(is, x)$, its characteristic function, is a continuous function of x at $s = 0$. By the continuity theorem for characteristic functions ([7], p. 30) the corresponding distribution function $Q(t, x)$ is therefore continuous in x for all $t > 0$. The continuity is uniform with respect to x in every continuity interval of t ([7], p. 31) and therefore $Q(t, x)$ is continuous in t and x combined for $t > 0$ and x in $B + C$, as well as for $t = 0$, $x \varepsilon B$. This proves the lemma.

Corollary. *The convergence of $P(t, x, \mu)$ to $Q(t, x)$, as $\mu \to 0$, is uniform for x in $B + C$ and $0 < t_0 \leq t \leq t_1$.* For, by a similar argument to that used in the preceding paragraph, it is seen that $P(t, x, \mu)$ is a continuous function of all arguments combined at $\mu = 0$, and this implies *uniform* continuity in the designated domain. This proves the last sentence of Theorem 5.

Lemma 5. *Let $u_k(x, \mu)$ satisfy the recursive relations*

$$\begin{aligned} u_{k+1}(x, \mu) &= \int_E u_k(y, \mu)\, dF(y, x, \mu) + \mu a_k(x, \mu), && x \varepsilon B, \\ u_k(x, \mu) &= b_k(x, \mu), && x \varepsilon E - B, \end{aligned} \tag{3.10}$$

and let $|a_k|$, $|b_k|$, *and* $|u_0|$ *be less than a constant M. Then, if Assumptions B, L, E, and S are satisfied, the inequality*

$$|u_k(x, \mu)| \leq C \cdot M$$

holds, where C is a constant independent of M.

PROOF. Assume, at first, that b_k and u_0 are identically zero, and that $a_k \equiv 1/\mu$. We denote the solution for this special case by u_k^*. It was proved in [1], Lemma 2, that $u_k^*(x)$ tends monotonically to the first moment $M_1(x)$ of N_x, as $k \to \infty$. From Theorem 4 we know that $\lim_{\mu \to 0} \mu M_1(x) = m_1(x)$, uniformly in B. Hence, $0 < u_k^* \leq C/\mu$.

Next, we drop the assumption $a_k \equiv 1/\mu$ and call the solution of the integral relation (3.10), in that case, $u_k^{(1)}$. Then the function $u_k^{**} = \mu M u_k^* - u_k^{(1)}$ solves the problem

$$u_{k+1}^{**} = \int_E u_k^{**}\, dF + \mu(M - a_k) \quad \text{in} \quad B,$$

$$u_k^{**} = 0 \quad \text{in} \quad E - B, \ u_0^* = 0 \quad \text{in} \quad B.$$

Since $M - a_k \geq 0$, it follows that $u_k^{**} \geq 0$ in B for all k, i.e., $u_1^{(1)} \leq$ const. M. The inequality $-u_1^{(1)} \leq$ const. M is proved analogously. Thus the lemma is proved in this special case.

Now we take the solution $u_k^{(2)}(x, \mu)$ of the special case that $a_k \equiv u_0 \equiv 0$. Here we obtain immediately by recursion the inequality $|u_k^{(2)}(x, \mu)| \leq M$. The solution $u_k^{(3)}(x, \mu)$ of the special case $a_k \equiv b_k \equiv 0$ also satisfies trivially the inequality $|u_k^{(3)}(x, \mu)| \leq M$.

Since the solution in the general case is the sum of three solutions corresponding to the three special cases, the lemma is proved.

PROOF OF THEOREM 5. Instead of comparing $P(t, x, \mu)$ directly with the solution of (3.5) we introduce the solution v of the problem

$$L[v] - \frac{\partial v}{\partial t} = 0 \quad \text{in} \quad B, \qquad t > 0, \qquad x \,\varepsilon\, B, \tag{3.11}$$

$$v(0, x) = Q(t_0, x), \qquad x \,\varepsilon\, B, \tag{3.12}$$

$$v(t, x) = 1, \qquad x \,\varepsilon\, E - B. \tag{3.13}$$

By this device we avoid difficulties connected with the discontinuity in the boundary conditions in (3.5) at $t = 0$, $x \,\varepsilon\, C$.

As in the proof of Theorem 3 we replace $v(t, x)$ by the function $v_\delta(t, x)$ defined by

$$v_\delta(t', x') = v(t, x), \qquad x' \,\varepsilon\, B'', \qquad t' > -\delta^3,$$

$$v_\delta(t, x) = 0, \qquad x \,\varepsilon\, E - B,$$

where x' is the function of x and δ introduced in Section 2, and

$$t' = \begin{cases} t & \text{for} \quad t > \delta, \\ t - (\delta - t)^3 & \text{for} \quad 0 \leq t \leq \delta. \end{cases} \tag{3.14}$$

This function $v_\delta(x, t)$ possesses continuous second derivatives with respect to the x_i and t for $x \varepsilon B''$ and $t > -\delta^3$. It is therefore possible, as in Section 2 and in [2], to apply Taylor's formula with quadratic terms to $v_\delta(t, y)$. An application of Assumptions B and L yields, similarly as in [1] and in the proof of Theorem 3,

$$\int_E v_\delta(t, y)\, dF(y, x, \mu) = v_\delta(t, x) + \mu L[v_\delta] + \mu g_1(t, x, \mu, \delta), \qquad x \varepsilon B, \tag{3.15}$$

where the function g_1 has the property that for every $\epsilon > 0$, $\delta > 0$, $t_1 > 0$, a $\mu_0 > 0$ can be found such that

$$| g_1(t, x, \mu, \delta)| \le \epsilon, \qquad x \varepsilon B + C, \qquad 0 \le t \le t_1, \qquad \mu \le \mu_0 . \tag{3.16}$$

Now by the definition of v_δ it is possible to choose δ so small, independently of the value of μ, that

$$L[v_\delta] = \frac{\partial v_\delta}{\partial t} + g_2(t, x, \delta),$$

where $g_2(t, x, \delta)$ satisfies the same inequality as $g_1(t, x, \mu, \delta)$. Hence

$$\mu L[v_\delta] = v_\delta(t + \mu, x) - v_\delta(t, x) + \mu g_3(t, x, \mu, \delta), \tag{3.17}$$

where, for a certain positive $\mu_1 \le \mu_0$, depending on δ and ϵ,

$$g_3(t, x, \mu, \delta) \le 2\epsilon, \qquad x \varepsilon B + C, \qquad 0 \le t \le t_1, \qquad \mu \le \mu_1 . \tag{3.18}$$

Combining (3.17) and (3.15) we find

$$v_\delta(t + \mu, x) = \int_E v_\delta(t, y)\, dF(y, x, \mu) + \mu h(t, x, \mu, \delta), \qquad x \varepsilon B, \tag{3.19}$$

where

$$| h(t, x, \mu, \delta)| \le 3\epsilon, \qquad x \varepsilon B + C, \qquad 0 \le t \le t_1, \qquad \mu \le \mu_1 . \tag{3.20}$$

Subtraction of (3.19) from (3.3) yields for

$$\omega(t, x, \mu) = P(t + t_0, x, \mu) - v_\delta(t, x) \tag{3.21}$$

the integral equation problem

$$\omega(t + \mu, x, \mu) = \int_E \omega(t, y, \mu)\, dF(y, x, \mu) + \mu h(t, x, \mu, \delta), \qquad x \varepsilon B, \quad t > 0, \tag{3.22}$$

$$\omega(t, x, \mu) = \omega_1(t, x, \mu), \qquad x \varepsilon E - B, \tag{3.23}$$

$$\omega(0, x, \mu) = \omega_2(x, \mu). \tag{3.24}$$

Here

$$| \omega_1(t, x, \mu)| = | P(t + t_0, x, \mu) - v_\delta(t, x)| = | 1 - v_\delta(t, x)| \le \epsilon \tag{3.25}$$

for $x \in E - B$, $0 \leq t \leq t_1$, provided δ has been chosen sufficiently small (independently of μ). This follows from (3.13) and the continuity properties of $v_\delta(t, x)$. Similarly we have, if δ and μ_1 are, independently of each other, chosen sufficiently small,

$$(3.26) \qquad \begin{aligned} | \omega_2(x, \mu)| = | P(t_0, x, \mu) - v_\delta(0, x)| &\leq | P(t_0, x, \mu) - Q(t_0, x)| \\ &+ | v(0, x) - v_\delta(0, x)| \leq \epsilon, \qquad x \in B + C. \end{aligned}$$

If we set $t = k\mu$ and write

$$\omega(k\mu, x, \mu) = u_k(x, \mu),$$

we can apply Lemma 5 to formulas (3.21) to (3.26) with the result that, if δ is sufficiently small,

$$| P(t + t_0, x, \mu) - v_\delta(t, x)| \leq 4C\epsilon,$$

for $x \in B + C$, $0 \leq t \leq t_1$, and $\mu \leq \mu_1$.

Finally, since v_δ differs arbitrarily little from v for sufficiently small δ, and ϵ was arbitrary, it follows that $Q(t, x) = \lim_{\mu \to 0} P(t, x, \mu)$ is, for all $x \in B$, the solution of the differential equation problem (3.11) to (3.13). By Lemma 4, $Q(t, x)$ approaches its values on C and its initial values for $t = 0$, as x approaches C or $t \to 0$, respectively, and is, therefore, indeed the solution of problem (3.5).

4. Some applications. If $L[u]$ is self-adjoint, then the solution of (2.6) can be calculated in the usual way by expansion in terms of the orthonormal eigenfunctions $u_j(x)$ of $L[u] + \lambda u = 0$, corresponding to the eigenvalues $\lambda = \lambda_j$, which are all real and positive. To do this we set $u = w + 1$ in (2.6) and solve the resulting problem

$$L[w] + sw = -s \text{ in } B, v = 0 \text{ on } C,$$

by the standard methods. (Cf., e.g., [9], p. 312. The argument for ordinary differential equations given there can be extended to partial differential equations whenever the existence of Green's function is known.) We find

$$(4.1) \qquad \begin{aligned} u(s, x) = 1 + w(s, x) &= 1 + \sum_{j=1}^{\infty} \frac{s}{\lambda_j - s} \int_B u_j(y)\, dy \cdot u_j(x) \\ &= 1 + \sum_{j=1}^{\infty} \left(\frac{\lambda_j}{\lambda_j - s} - 1 \right) \int_B u_j(y)\, dy \cdot u_j(x). \end{aligned}$$

The series $\sum_{j=1}^{\infty} \int_B u_j(y)\, dy \cdot u_j(x)$ is the generalized Fourier series of the function that is identically one in B. If this series actually converges to 1 in the interior of B, formula (4.1) simplifies to

$$(4.2) \qquad u(s, x) = \sum_{j=1}^{\infty} \frac{\lambda_j}{\lambda_j - s} \int_B u_j(y)\, dy \cdot u_j(x), \qquad x \in B.$$

From here on we assume explicitly that (4.2) is valid:

ASSUMPTION C. *The series* $\sum_{j=1}^{\infty} \int_B u_j(y)\, dy \cdot u_j(x)$ *converges for all* $x \varepsilon B$.

In this case we can give an explicit expression for the distribution function $Q(t, x)$, for

$$u(s, x) = \int_0^{\infty} e^{st} \sum_{j=1}^{\infty} \lambda_j \int_B u_j(y)\, dy \cdot u_j(x) e^{-\lambda_j t}\, dt,$$

and therefore, because of (3.4),

$$Q(t, x) = 1 - \sum_{j=0}^{\infty} \int_B u_j(y)\, dy \cdot u_j(x) e^{-\lambda_j t}. \tag{4.3}$$

This proves

THEOREM 6. *If the Assumptions B, L, E, S, and C are satisfied, if* $L[u]$ *is self-adjoint, and if the lowest eigenvalue* λ_1 *of* $L[u] + \lambda u = 0$ *is simple, then*

$$Pr\{N_x \geq k\} = u_1(x) \int_B u_1(y)\, dy \cdot e^{-\lambda_1 k\mu} + O(e^{-(\lambda_1 - \lambda_2)k\mu}) + \alpha(k\mu, x, \mu), \tag{4.4}$$

where

$$\lim_{\mu \to 0} \alpha(t, x, \mu) = 0$$

uniformly in t and x.

The leading term in (4.4) is thus a good approximation to $Pr\{N_x \geq k\}$ in a range of the variables μk for which $k\mu$ is so large, and at the same time μ so small, that the two remainder terms can be neglected.

The preceding calculations have some points of contact with those of M. Kac in [10], Section 10. The results there refer to the special case of Brownian motion. Also, an integral equation is used instead of (2.6), which permits a considerable relaxation of the condition S.

As a special application we consider random walks for which $L[u]$ reduces to a constant multiple of the Laplacian. It can then be assumed without loss of generality (see [1], Section 4) that μ is the mean square of the step length and that

$$L[u] = \frac{1}{2n} \Delta u. \tag{4.5}$$

A domain B for which all quantities involved can be calculated explicitly is the n-dimensional sphere of radius a with center at $x = 0$. A routine calculation leads to the formula

$$u(s, x) = \left(\frac{r}{a}\right)^{1-n/2} J_{n/2-1}(\sqrt{2ns}r)/J_{n/2-1}(\sqrt{2ns}a) \tag{4.6}$$

for the moment generating function. Here $r = (\sum_{j=1}^{n} x_j^2)^{\frac{1}{2}}$ and $J_k(z)$ is Bessel's

function of the first kind. The series for $u(s, x)$ in powers of s,

$$u(s, x) = 1 + (a^2 - r^2)s + \frac{a^2 - r^2}{2n + 4}[(n + 4)a^2 - nr^2]s^2 + \cdots,$$

and an application of Theorems 3 and 4 lead to the expressions

$$\lim_{\mu \to 0} \mu E[N_x] = a^2 - r^2,$$

$$\lim_{\mu \to 0} \mu^2 E[N_x - E[N_x]] = \frac{2}{n + 2}(a^4 - r^4)$$

for the mean and the variance of the duration. The relative error $e[N_x]$, i.e., the standard deviation of N_x divided by its mean value, satisfies therefore the relation

$$\lim_{\mu \to 0} e[N_x] = \sqrt{\frac{2}{n + 2}} \sqrt{\frac{a^2 + r^2}{a^2 - r^2}}. \tag{4.7}$$

It should be noted that *the relative error is a decreasing function of the number of dimensions.*

We omit the straightforward calculation needed for the determination of the eigenfunctions $u_j(x)$ and eigenvalues λ_j in the present case and state only the results:

Let $\rho = \rho_j$ be the jth positive zero, in order of increasing size, of the function $J_{n/2-1}(\rho)$; then

$$\lambda_j = \rho_j^2/2na^2.$$

Assumption C is satisfied (cf., e.g., [11], p. 591) and

$$Q(t, x) = 1 - 2\left(\frac{r}{a}\right)^{1-n/2} \sum_{j=1}^{\infty} \frac{J_{n/2-1}(\rho_j r/a)}{\rho_j J_{n/2}(\rho_j)} e^{-\rho_j^2 t/2na^2}. \tag{4.8}$$

For $n = 2$, we find, e. g., using the approximation (4.4), for small μ, and Theorem 4 with $k = 1$,

$$Pr\{N_0 \geq 2E[N_0]\} \sim 1/11. \tag{4.9}$$

REFERENCES

[1] W. WASOW, "On the mean duration of random walks," *Jour. Research Nat. Bur. Standards*, Vol. 46 (1951).

[2] I. PETROWSKY, "Ueber das Irrfahrtproblem," *Math. Ann.*, Vol. 109 (1934), pp. 425–444.

[3] J. H. CURTISS, "On the theory of moment generating functions," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 430–433.

[4] A. KHINTCHINE, *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*, Julius Springer, Berlin, 1933.

[5] P. ERDÖS AND M. KAC, "On the number of positive sums of independent random variables," *Bull. Am. Math. Soc.*, Vol. 53 (1947), pp. 1011–1020.

[6] F. G. DRESSEL, "The fundamental solution of the parabolic equation II," *Duke Math. Jour.*, Vol. 13 (1946), pp. 61–70.

[7] H. Cramér, *Random Variables and Probability Distributions*, Cambridge University Press, 1937.
[8] M. Gevrey, "Sur les équations aux dérivées partielles du type parabolique," *Jour. Math. Pures Appl.*, Vol. 9 (1913), pp. 305–471.
[9] R. Courant and D. Hilbert, *Methoden der Mathematischen Physik*, Vol. 1, 2nd ed., Julius Springer, Berlin, 1931.
[10] M. Kac, "On some connections between probability theory and differential and integral equations," to be published.
[11] G. N. Watson, *Theory of Bessel Functions*, 2nd ed., Cambridge University Press, 1948.
[12] W. Feller, "Zur Theorie der stochastischen Prozesse," *Math. Ann.*, Vol. 113 (1937), pp. 113–161.
[13] R. Fortet, "Les fonctions aléatoires du type de Markoff associées à certaines équations linéaires aux dérivées partielles du type parabolique," *Jour. Math. Pures Appl.*, 9th series, Vol. 22 (1943), pp. 177–244.

ON THE THEORY OF UNBIASED TESTS OF SIMPLE STATISTICAL HYPOTHESES SPECIFYING THE VALUES OF TWO OR MORE PARAMETERS[1]

By Stanley L. Isaacson

Columbia University and Iowa State College

1. Summary. Unbiased critical regions of type D for testing simple hypotheses specifying the values of several parameters are defined and their properties studied. These regions constitute a natural generalization of the Neyman-Pearson regions of type A for testing simple hypotheses specifying the value of one parameter. A theorem is obtained which plays the role of the Neyman-Pearson fundamental lemma in the type A case. Illustrative examples of type D regions are given.

2. Introduction. The parameter space Ω will, in this paper, be a subset of a k-dimensional Euclidean space ($k \geq 1$), and $\theta = (\theta_1, \cdots, \theta_k)$ will denote a point in Ω. A simple statistical hypothesis is one which specifies the values of all unknown parameters. When we refer to a statistical test we mean a Borel measurable set in an n-dimensional sample space such that if the sample point falls in this critical region we reject the null hypothesis. In this paper the term *region* will always mean a Borel measurable set. The probability of rejecting a true hypothesis when using a given test is called the size of this test. A test is unbiased if the power function of the test has a relative minimum for the value $\theta = \theta^0$, where θ^0 is the value of θ specified by the hypothesis to be tested.

A locally best unbiased region for testing a simple hypothesis specifying the value of one parameter is called *type A* by Neyman and Pearson [1]. It is obtained by maximizing the curvature of the power curve at the point $\theta = \theta^0$ specified by the hypothesis, subject to the conditions of given size and unbiasedness. Geometrically speaking, the power curve of a region of type A is above the power curves of all other unbiased regions of the same size in an infinitesimal neighborhood of θ^0. For the purpose of generalization to the k-parameter case it is useful to note that if we consider the power curve of the type A region and the power curves of any other unbiased regions of the same size, then the length of a horizontal chord at a fixed infinitesimal distance above the minimum point is a minimum when compared with the length of this chord on the power curves of the other unbiased regions of the same size. We note that the definition of type A regions does not use any information about the relative importance of errors of type II. (An error of type II is made when we accept a false hypothesis.)

Type A regions remain invariant under transformations of the parameter which are locally one-to-one and twice differentiable. Regions of type A can

---

[1] Work done under the sponsorship of the Office of Naval Research.

be proved to exist under quite weak assumptions on the joint density function of the sample.

When we proceed to consider simple hypotheses specifying the values of two or more parameters, we are immediately faced with a more complicated situation. (For the sake of simplicity in statement, we confine ourselves in this introductory section to the two-parameter theory; the extension of our discussion to three or more parameters is direct.) In the two-parameter case the geometrical picture of the power function is a surface, and if we require of a locally best unbiased region that its power surface have maximum curvature along every cross-section at the point $(\theta_1, \theta_2) = (\theta_1^0, \theta_2^0)$ specified by the hypothesis, subject to the conditions of size and unbiasedness, then it develops that this requirement cannot be met even in the simplest cases; for if we maximize the curvature of the power surface along one of its cross-sections, we find that in general this causes the curvature to diminish along other cross-sections and so we cannot maximize the curvature along all cross-sections at once.

To handle the two-parameter theory, Neyman and Pearson [2] considered *type C* regions. They require of a critical region not only that it be of given size and unbiased but also that it have constant power in an infinitesimal neighborhood of (θ_1^0, θ_2^0) along a *given* family of concentric ellipses with the same shape and orientation; the type C region is then defined as the one among this class of regions which gives best local power. When the *given* family of ellipses consists of circles, the region of type C is called *regular*; otherwise it is called *nonregular*. One can choose the family of ellipses if and essentially only if one knows the relative importance of errors of type II in an infinitesimal neighborhood of (θ_1^0, θ_2^0). In the absence of such information one cannot proceed to find a region of type C. Regions of type C retain their property of unbiasedness under transformations of the parameter space which are locally one-to-one and twice differentiable, but in general *regular* unbiased critical regions of type C become *nonregular* under such transformations. Hence if one is inclined in the absence of advance information about errors of type II to favor the regular unbiased region of type C as a region fulfilling "good" intuitive requirements, then the objection can be raised that these regular regions of type C are not invariant under transformations of the parameter space.

There is an approach to the problem of finding a "good" critical region which overcomes the objections raised to the type C theory; i.e., it will provide us with a criterion for choosing a critical region without using any advance knowledge as to the relative importance of errors of type II, and this type of critical region will be invariant under transformations of the parameters. This type of critical region, which will be a natural generalization of the type A region of the one-parameter theory, will maximize the Gaussian curvature of the power surface at $(\theta_1, \theta_2) = (\theta_1^0, \theta_2^0)$, subject to the conditions of size and unbiasedness. In the next two sections we shall develop this theory for simple hypotheses specifying the values of two parameters, and then in Section 5 we shall extend it to the case of simple hypotheses specifying the values of three or more parameters.

3. Definition of unbiased critical regions of type D in the two-parameter case. We introduce for the power function of a region w the symbol

$$\beta(\theta_1, \theta_2 \mid w) = Pr(E \,\varepsilon\, w \mid \theta_1, \theta_2), \tag{3.1}$$

where $E = (x_1, \cdots, x_n)$ is the sample point in an n-dimensional sample space. Here the joint probability distribution of the sample depends on the parameter $\theta = (\theta_1, \theta_2)$, and we are testing the hypothesis $(\theta_1, \theta_2) = (\theta_1^0, \theta_2^0)$. We make a translation of the parameter space to bring the point (θ_1^0, θ_2^0) to the origin, so that we may consider the test of the hypothesis $(\theta_1, \theta_2) = (0, 0)$. The size of the critical region is then

$$\beta(0, 0 \mid w) = Pr(E \,\varepsilon\, w \mid 0, 0). \tag{3.2}$$

We also introduce the following notation:

$$\beta_i(w) = \left.\frac{\partial\beta(\theta_1, \theta_2 \mid w)}{\partial\theta_i}\right|_{(\theta_1,\theta_2)=(0,0)}, \qquad i = 1, 2, \tag{3.3}$$

$$\beta_{ij}(w) = \left.\frac{\partial^2\beta(\theta_1, \theta_2 \mid w)}{\partial\theta_i\,\partial\theta_j}\right|_{(\theta_1,\theta_2)=(0,0)}, \qquad i, j = 1, 2. \tag{3.4}$$

We assume these derivatives exist. We shall write β_i and β_{ij} for $\beta_i(w)$ and $\beta_{ij}(w)$, respectively, whenever our doing so will cause no ambiguity. We note that the derivatives are taken at $(\theta_1, \theta_2) = (0, 0)$, though this fact does not show up in our notation.

In books on differential geometry, such as Eisenhart [3], it is shown that if we consider a surface in three-dimensional Euclidean space and a point (x_0, y_0) at which the second partial derivatives of the function $z = f(x, y)$ which describes the surface exist and are continuous, then the so-called Gaussian or total curvature K of the surface $z = f(x, y)$ at the point (x_0, y_0) is given by:

$$K = \frac{\left.\dfrac{\partial^2 z}{\partial x^2}\right|_{(x_0,y_0)} \left.\dfrac{\partial^2 z}{\partial y^2}\right|_{(x_0,y_0)} - \left[\left.\dfrac{\partial^2 z}{\partial x\partial y}\right|_{(x_0,y_0)}\right]^2}{\left(1 + \left[\left.\dfrac{\partial z}{\partial x}\right|_{(x_0,y_0)}\right]^2 + \left[\left.\dfrac{\partial z}{\partial y}\right|_{(x_0,y_0)}\right]^2\right)^2}. \tag{3.5}$$

The Gaussian curvature is invariant under translation and rotation of the coordinate axes. Applying (3.5) to the power surface $\beta = \beta(\theta_1, \theta_2 \mid w)$ at the point $(\theta_1, \theta_2) = (0, 0)$, imposing the condition of unbiasedness on w, and noting that necessary conditions for an unbiased region are that $\beta_1(w) = \beta_2(w) = 0$, we have

$$K = \frac{\beta_{11}(w)\beta_{22}(w) - \beta_{12}^2(w)}{(1 + 0 + 0)^2} = \begin{vmatrix} \beta_{11}(w) & \beta_{12}(w) \\ \beta_{12}(w) & \beta_{22}(w) \end{vmatrix} = \det B_w, \tag{3.6}$$

where

$$B_w = \begin{pmatrix} \beta_{11}(w) & \beta_{12}(w) \\ \beta_{12}(w) & \beta_{22}(w) \end{pmatrix}.$$

As a natural generalization of the type A region of the one-parameter theory, we now propose as a critical region for testing $(\theta_1, \theta_2) = (0, 0)$ that critical region which maximizes the Gaussian curvature of the power surface at (0, 0), subject to the conditions of size and unbiasedness. This leads us to the following

DEFINITION. *A region w_0 is said to be an unbiased critical region of type D for testing H_0 if:*

I. $\beta(0, 0 \mid w_0) = \alpha$;

II. $\beta_i(w_0) = 0, \qquad i = 1, 2$;

III. *B_{w_0} is positive definite*;

IV. *$\det B_{w_0} \geq \det B_w$ for any other region w satisfying* I–III.

Condition I specifies the size of the test. Conditions II and III insure the existence of a relative minimum at $(\theta_1, \theta_2) = (0, 0)$ and so imply the condition of unbiasedness. Condition IV specifies that the region of type D has maximal Gaussian curvature among all unbiased regions of the prescribed size.

Let us consider the geometrical interpretation of a region of type D. In the one-parameter theory we noted that the type A region minimizes the length of a certain infinitesimal chord on the power curve. We shall now see that the type D region minimizes the area of a certain infinitesimal ellipse, subject to the conditions of size and unbiasedness. Consider a Taylor expansion of the power function in an infinitesimal neighborhood of $(\theta_1, \theta_2) = (0, 0)$. We have, neglecting infinitesimals of the third and higher orders,

$$\begin{aligned}\beta(\theta_1, \theta_2 \mid w) &= \beta(0, 0 \mid w) + \theta_1\beta_1 + \theta_2\beta_2 + \tfrac{1}{2}(\theta_1^2\beta_{11} + 2\theta_1\theta_2\beta_{12} + \theta_2^2\beta_{22}) \\ &= \alpha + \tfrac{1}{2}(\theta_1^2\beta_{11} + 2\theta_1\theta_2\beta_{12} + \theta_2^2\beta_{22}).\end{aligned} \tag{3.7}$$

Consider the ellipse $\theta_1^2\beta_{11} + 2\theta_1\theta_2\beta_{12} + \theta_2^2\beta_{22} = \delta$, where δ is a positive constant; this ellipse is a horizontal cross-section of the power surface at an infinitesimal distance above the minimum point $(\theta_1, \theta_2) = (0, 0)$. It is well known that the area of this ellipse is given by

$$\frac{\pi\delta}{\sqrt{\begin{vmatrix}\beta_{11} & \beta_{12} \\ \beta_{12} & \beta_{22}\end{vmatrix}}} = \frac{\pi\delta}{\sqrt{\det B}}. \tag{3.8}$$

We have just seen that the region of type D maximizes the determinant of B subject to the conditions of size and unbiasedness. Hence it minimizes the area of our infinitesimal ellipse subject to these same conditions.

4. Theorems concerning regions of type D in the two-parameter theory. Having defined regions of type D, we now wish to obtain a theorem which will characterize the structure of such regions for us. We shall assume the following fundamental condition is satisfied:

There exists a joint density function $p(E \mid \theta_1, \theta_2)$ for any point (θ_1, θ_2) in the parameter space Ω; and for any fixed region w in the sample space the integral $\int_w p(E \mid \theta_1, \theta_2)\, dE$ has second partial derivatives with respect to θ_1 and θ_2 in a

neighborhood of $(\theta_1, \theta_2) = (0, 0)$ *which are continuous at* $(0, 0)$, *and the integral can be differentiated twice under the integral sign with respect to* θ_1 *and* θ_2 *at* $(0, 0)$.

The derivatives of the above types taken at $(\theta_1, \theta_2) = (0, 0)$ will be denoted simply as follows:

$$\frac{\partial}{\partial\theta_i}\int_w p(E \mid \theta_1, \theta_2)\, dE \,\Big|_{\theta_1=\theta_2=0} = \int_w p_i\, dE = \beta_i(w), \qquad i = 1, 2, \tag{4.1}$$

$$\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\int_w p(E \mid \theta_1, \theta_2)\, dE \,\Big|_{\theta_1=\theta_2=0} = \int_w p_{ij}\, dE = \beta_{ij}(w), \qquad i, j = 1, 2, \tag{4.2}$$

where

$$p_i = \frac{\partial p(E \mid \theta_1, \theta_2)}{\partial\theta_i}\Bigg|_{\theta_1=\theta_2=0}, \qquad p_{ij} = \frac{\partial^2 p(E \mid \theta_1, \theta_2)}{\partial\theta_i\,\partial\theta_j}\Bigg|_{\theta_1=\theta_2=0}. \tag{4.3}$$

We also write $p(E \mid 0, 0) = p$.

We seek a theorem which will tell us how to characterize the structure of a region w_0 such that $\int_w g_{11}\, dE \int_w g_{22}\, dE - \left(\int_w g_{12}\, dE\right)^2$ is a maximum, subject to the side conditions that $\int_w f_i\, dE = c_i$, $i = 1, \cdots, m$, where the g_{ij} and the f_i are given integrable functions and the c_i are given constants. If we have such a theorem, then by taking $m = 3$, $g_{11} = p_{11}$, $g_{12} = p_{12}$, $g_{22} = p_{22}$, $f_1 = p$, $f_2 = p_1$, $f_3 = p_2$, $c_1 = \alpha$, $c_2 = 0$, $c_3 = 0$, we will be able to use the theorem to characterize the structure of a region of type D, since in terms of the p's our conditions on a type D region are:

I′. $\int_{w_0} p\, dE = \alpha$;

II′. $\int_{w_0} p_i\, dE = 0, \qquad i = 1, 2;$

III′. *The matrix* $P_{w_0} = \left(\int_{w_0} p_{ij}\, dE\right)$, $i, j = 1, 2$, *is positive definite*;

IV′. *det* $P_{w_0} \geq$ *det* P_w *for any other region* w *satisfying* I′–III′.

We will now state the Neyman-Pearson fundamental lemma, which is used in the one-parameter theory to find regions of type A, in order to indicate the type of theorem we are seeking and also because we shall use this lemma in proving our theorem.

THE NEYMAN-PEARSON FUNDAMENTAL LEMMA. *Suppose* $m + 1$ *given integrable functions* $f_0, f_1, \cdots, f_m$ *are defined in an n-dimensional space. Consider the set of all regions* w *for which the following conditions are fulfilled*:

$$\int_w f_i\, dE = c_i, \qquad i = 1, \cdots, m, \tag{4.4}$$

where the c_i are m given constants. If w_0 is a region which satisfies the m conditions (4.4) *and if*

$$f_0 \geq \sum_{i=1}^{m} k_i f_i \quad in \quad w_0, \qquad (4.5)$$
$$f_0 \leq \sum_{i=1}^{m} k_i f_i \quad outside \quad w_0,$$

for m suitably chosen constants k_i, then w_0 has the property that

$$\int_{w_0} f_0 \, dE \geq \int_w f_0 \, dE \qquad (4.6)$$

for any region w which satisfies (4.4).

We proceed to state and prove a lemma which will tell us how to characterize a region w_0 maximizing $\int_w g_{11} \, dE \int_w g_{22} \, dE$ subject to integral side conditions, and then to use the lemma and a corollary to it to prove a theorem which will characterize the structure of a region w_0 which maximizes $\int_w g_{11} \, dE \int_w g_{22} \, dE - \left(\int_w g_{12} \, dE\right)^2$ subject to integral side conditions.

LEMMA 1. *Suppose $m + 2$ given integrable functions $g_{11}, g_{22}, f_1, \cdots, f_m$ are defined in an n-dimensional space. Consider the set of all regions w for which the following conditions are fulfilled:*

$$\int_w f_i \, dE = c_i, \qquad i = 1, \cdots, m, \qquad (4.7)$$

$$\int_w g_{jj} \, dE > 0, \qquad j = 1, 2, \qquad (4.8)$$

where the c_i are m given constants. If w_0 is a region which satisfies conditions (4.7) *and* (4.8), *and if*

$$\sum_{i=1}^{2} k_{ii} g_{ii} \geqq \sum_{i=1}^{m} k_i f_i \quad in \quad w_0, \qquad (4.9)$$
$$\sum_{i=1}^{2} k_{ii} g_{ii} \leqq \sum_{i=1}^{m} k_i f_i \quad outside \quad w_0,$$

where $k_{11} = \int_{w_0} g_{22} \, dE$, $k_{22} = \int_{w_0} g_{11} \, dE$, and the k_i are m suitably chosen constants, then w_0 has the property that

$$\prod_{j=1}^{2} \int_{w_0} g_{jj} \, dE \geqq \prod_{j=1}^{2} \int_w g_{jj} \, dE \qquad (4.10)$$

for any region w which satisfies (4.7) *and* (4.8).

We note that we must know our region w_0 in advance so that we can calculate k_{11} and k_{22} and thus verify whether w_0 has the structure required by the lemma or not.

PROOF. We apply the Neyman-Pearson fundamental lemma to the function

$$f_0 = \sum_{i=1}^{2} k_{ii} g_{ii} = g_{11} \int_{w_0} g_{22}\, dE + g_{22} \int_{w_0} g_{11}\, dE.$$

From (4.6) we obtain

$$\int_{w} g_{11}\, dE \int_{w_0} g_{22}\, dE + \int_{w} g_{22}\, dE \int_{w_0} g_{11}\, dE \leq 2 \int_{w_0} g_{11}\, dE \int_{w_0} g_{22}\, dE \tag{4.11}$$

for any region w satisfying (4.7). Knowing (4.11) we must prove that

$$\int_{w} g_{11}\, dE \int_{w} g_{22}\, dE \leq \int_{w_0} g_{11}\, dE \int_{w_0} g_{22}\, dE \tag{4.12}$$

for any region w satisfying (4.7) and (4.8).

Let

$$x_j = \frac{\int_{w} g_{jj}\, dE}{\int_{w_0} g_{jj}\, dE}, \qquad j = 1, 2. \tag{4.13}$$

Since the integrals $\int_{w} g_{jj}\, dE$, $\int_{w_0} g_{jj}\, dE$, $j = 1, 2$, are all positive by (4.8) we may rewrite (4.11) and (4.12) in terms of the x_j's as follows:

$$x_1 + x_2 \leq 2, \tag{4.14}$$

$$x_1 x_2 \leq 1. \tag{4.15}$$

Thus we must prove that $x_1 x_2 \leq 1$ whenever $\frac{1}{2}(x_1 + x_2) \leq 1$, where x_1 and x_2 are positive real numbers. But this follows immediately from the well known inequality between the arithmetic and geometric mean, and hence our lemma is proved.

COROLLARY. *If a region w_0 satisfies conditions (4.7), (4.8), and (4.9) of the lemma, and if g_{12} is a given integrable function for which $\int_{w_0} g_{12}\, dE = 0$, then*

$$\begin{aligned} &\int_{w_0} g_{11}\, dE \int_{w_0} g_{22}\, dE - \left(\int_{w_0} g_{12}\, dE\right)^2 \\ &\qquad \geq \int_{w} g_{11}\, dE \int_{w} g_{22}\, dE - \left(\int_{w} g_{12}\, dE\right)^2 \end{aligned} \tag{4.16}$$

for any region w satisfying conditions (4.7) and (4.8) of the lemma.

We now use the lemma and the corollary to it to prove the following theorem:

THEOREM 1. *Suppose the elements g_{ij} of a symmetric 2×2 matrix*

$$G = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}$$

are given integrable functions defined in an n-dimensional space; and that $f_1, \cdots, f_m$ are m other given integrable functions defined in this space. For any region w, let

$$G_w = \begin{bmatrix} \int_w g_{11}\, dE & \int_w g_{12}\, dE \\ \int_w g_{21}\, dE & \int_w g_{22}\, dE \end{bmatrix}.$$

Consider the set of all regions w for which the following conditions are fulfilled:

$$\int_w f_i\, dE = c_i, \qquad i = 1, \cdots, m, \tag{4.17}$$

$$G_w \text{ is positive definite,} \tag{4.18}$$

where the c_i are m given constants. If w_0 is a region which satisfies the conditions (4.17) *and* (4.18), *and if*

$$\begin{aligned} \sum_{i,j=1}^{2} k_{ij} g_{ij} &\geq \sum_{i=1}^{m} k_i f_i \quad \text{in} \quad w_0, \\ \sum_{i,j=1}^{2} k_{ij} g_{ij} &\leq \sum_{i=1}^{m} k_i f_i \quad \text{outside} \quad w_0, \end{aligned} \tag{4.19}$$

where $k_{11} = \int_{w_0} g_{22}\, dE$, $k_{22} = \int_{w_0} g_{11}\, dE$, $k_{12} = k_{21} = -\int_{w_0} g_{12}\, dE$, and the k_i are m suitably chosen constants, then w_0 has the property that

$$\det G_{w_0} \geq \det G_w \tag{4.20}$$

for any region w which satisfies (4.17) *and* (4.18).

PROOF. We know there exists an orthogonal matrix H of constants which diagonalizes G_{w_0}; that is, $H'G_{w_0}H$ is a diagonal matrix, and $H'H = I$. Apply this transformation to

$$G = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}$$

and let

$$G^* = \begin{pmatrix} g_{11}^* & g_{12}^* \\ g_{21}^* & g_{22}^* \end{pmatrix} = H'GH.$$

We note that

$$
(4.21) \qquad G^*_{w_0} = \begin{pmatrix} \int_{w_0} g^*_{11}\, dE & 0 \\ 0 & \int_{w_0} g^*_{22}\, dE \end{pmatrix} = H'G_{w_0}H.
$$

Since H is orthogonal, we know that det G_{w_0} = det $G^*_{w_0}$, and also det G_w = det G^*_w, where

$$
G^*_w = \begin{pmatrix} \int_w g^*_{11}\, dE & \int_w g^*_{12}\, dE \\ \int_w g^*_{21}\, dE & \int_w g^*_{22}\, dE \end{pmatrix} = H'G_wH.
$$

Thus we see that if det $G^*_{w_0} \geq$ det G^*_w for any region w satisfying (4.17) and (4.18), then det $G_{w_0} \geq$ det G_w for any such region, and this is what we seek to prove. But since G_{w_0} and G_w are positive definite, we know that $G^*_{w_0}$ and G^*_w are positive definite; hence their diagonal elements are positive and they satisfy condition (4.8); then by our lemma and its corollary we know that det $G^*_{w_0} \geq$ det G^*_w for any region w satisfying (4.17) and (4.18) (and hence (4.8)), if w_0 satisfies

$$
(4.22) \qquad \begin{aligned} g^*_{11} \int_{w_0} g^*_{22}\, dE + g^*_{22} \int_{w_0} g^*_{11}\, dE &\geq \sum_{i=1}^{m} k_i f_i \quad \text{in} \quad w_0, \\ g^*_{11} \int_{w_0} g^*_{22}\, dE + g^*_{22} \int_{w_0} g^*_{11}\, dE &\leq \sum_{i=1}^{m} k_i f_i \quad \text{outside} \quad w_0. \end{aligned}
$$

It now remains only to prove that the conditions (4.22) are implied by (4.19). To do this we shall prove that

$$
g^*_{11} \int_{w_0} g^*_{22}\, dE + g^*_{22} \int_{w_0} g^*_{11}\, dE = g_{11} \int_{w_0} g_{22}\, dE + g_{22} \int_{w_0} g_{11}\, dE - 2g_{12} \int_{w_0} g_{12}\, dE.
$$

Denote the adjoint of a matrix A by adj A. Then (adj $G^*_{w_0})G^*$ = H'(adj $G_{w_0})HH'GH$ = H'(adj $G_{w_0})GH$, since H is orthogonal. But

$$
g_{11} \int_{w_0} g_{22}\, dE + g_{22} \int_{w_0} g_{11}\, dE - 2g_{12} \int_{w_0} g_{12}\, dE
$$

is the trace of (adj $G_{w_0})G$ and similarly

$$
g^*_{11} \int_{w_0} g^*_{22}\, dE + g^*_{22} \int_{w_0} g^*_{11}\, dE
$$

is the trace of (adj $G^*_{w_0})G^*$. Hence our two expressions are equal, as we know that the trace of a matrix is invariant under an orthogonal transformation, and we have just seen that (adj $G^*_{w_0})G^*$ is obtained from (adj $G_{w_0})G$ by such a transformation. This completes the proof.

We shall now prove a result we mentioned earlier; namely, the invariance of regions of type D under transformations of the parameters.

THEOREM 2. *If the transformation* $\theta_i = T_i(\Theta_1, \Theta_2)$, $i = 1, 2$, *is such that the first and second partial derivatives* $\partial\theta_s/\partial\Theta_i$ *and* $\partial^2\theta_s/\partial\Theta_i\partial\Theta_j$ *exist and are continuous at* $(\Theta_1, \Theta_2) = (0, 0)$, $i, j, s = 1, 2$, *the Jacobian* $\partial(\theta_1, \theta_2)/\partial(\Theta_1, \Theta_2)$ *differs from zero at* $(\Theta_1, \Theta_2) = (0, 0)$, *and* $(0, 0)$ *maps into* $(0, 0)$; *then a region* w_0, *which is an unbiased critical region of type D for testing* $(\theta_1, \theta_2) = (0, 0)$ *against the set of alternative hypotheses specifying the values of the parameters* θ_1 *and* θ_2, *will remain an unbiased critical region of type D for testing* $(\Theta_1, \Theta_2) = (0, 0)$ *against the set of transformed hypotheses specifying the values of the new parameters* Θ_1 *and* Θ_2.

PROOF. We adopt the following notation:

$$\left.\frac{\partial\theta_1}{\partial\Theta_1}\right|_{\Theta_1=\Theta_2=0} = K, \qquad \left.\frac{\partial\theta_1}{\partial\Theta_2}\right|_{\Theta_1=\Theta_2=0} = L,$$
$$\left.\frac{\partial\theta_2}{\partial\Theta_1}\right|_{\Theta_1=\Theta_2=0} = M, \qquad \left.\frac{\partial\theta_2}{\partial\Theta_2}\right|_{\Theta_1=\Theta_2=0} = N. \tag{4.23}$$

By hypothesis the determinant of

$$J = \begin{pmatrix} K & L \\ M & N \end{pmatrix}$$

is not equal to zero. We denote by $\beta_{(i)}$ and $\beta_{(ij)}$ the partial derivatives of the power function with respect to Θ_i and Θ_j *evaluated at* $(\Theta_1, \Theta_2) = (0, 0)$. Also let

$$B_{(w)} = \begin{pmatrix} \beta_{(11)}(w) & \beta_{(12)}(w) \\ \beta_{(12)}(w) & \beta_{(22)}(w) \end{pmatrix}.$$

Then we can write

$$\beta_{(1)} = \beta_1 K + \beta_2 M,$$
$$\beta_{(2)} = \beta_1 L + \beta_2 N. \tag{4.24}$$

The condition that $\beta(0, 0 \mid w_0) = \alpha$ is unchanged by the transformation of parameters. Since we know that for an unbiased region $\beta_1 = 0$ and $\beta_2 = 0$, we obtain from (4.24) that $\beta_{(1)} = 0$ and $\beta_{(2)} = 0$. Thus, since the partial derivatives of the transformation are continuous, our property of unbiasedness is retained. Also since $\beta_1 = 0$ and $\beta_2 = 0$, it is easily seen that

$$B_{(w)} = J' B_w J. \tag{4.25}$$

Since J is nonsingular by hypothesis, we know that $B_{(w)}$ is positive definite since B_w is a positive definite matrix. Also we have that

$$\det B_{(w)} = (\det J)^2 \det B_w\,; \tag{4.26}$$

and since $\det J \neq 0$, it follows that if $\det B_{w_0} \geq \det B_w$, then $\det B_{(w_0)} \geq$

det $B_{(w)}$. Thus we have seen that w_0 satisfies all the conditions of a region of type D for testing $(\Theta_1, \Theta_2) = (0, 0)$, and our proof is completed.

The inequalities which must hold within and outside the unbiased critical regions of type D can frequently be simplified if we express them in terms of the derivatives of $\log p(E \mid \theta_1, \theta_2)$. We write

$$(4.27)\qquad \begin{aligned} \phi_t &= \left.\frac{\partial \log p(E \mid \theta_1, \theta_2)}{\partial \theta_t}\right|_{(\theta_1,\theta_2)=(0,0)}, \\ \phi_{ts} &= \left.\frac{\partial^2 \log p(E \mid \theta_1, \theta_2)}{\partial \theta_t \partial \theta_s}\right|_{(\theta_1,\theta_2)=(0,0)}, \end{aligned}$$

where $t, s = 1, 2$. In particular, the simplification will be considerable if

$$(4.28)\qquad \phi_{ts} = A_{ts} + B_{ts}\phi_1 + C_{ts}\phi_2, \qquad t, s = 1, 2,$$

where A_{ts}, B_{ts}, C_{ts} are independent of the sample point E but may depend on (θ_1, θ_2). If (4.28) is true, it will be seen that

$$(4.29)\qquad p_1 = \phi_1 p, \qquad p_2 = \phi_2 p,$$

$$(4.30)\qquad p_{ts} = (\phi_t\phi_s + A_{ts} + B_{ts}\phi_1 + C_{ts}\phi_2)p.$$

Consequently, the type of inequalities (4.19) occurring among the sufficient conditions of Theorem 1 for a region of type D will reduce to the following for points where $p > 0$ (assuming that $W_+(\theta_1, \theta_2) = \{E \mid p(E \mid \theta_1, \theta_2) > 0\}$ is independent of (θ_1, θ_2)):

$$(4.31)\qquad \left(\int_{w_0} p_{11}\, dE\right)\phi_2^2 + \left(\int_{w_0} p_{22}\, dE\right)\phi_1^2 - 2\left(\int_{w_0} p_{12}\, dE\right)\phi_1\phi_2 \gtrless k_1' + k_2'\phi_1 + k_3'\phi_2,$$

where the k_i' are new constants easily expressible in terms of the k_i, $\int_{w_0} p_{ij}\, dE$, and the coefficients in (4.28). The k_i' must be determined so as to satisfy $\int_{w_0} p\, dE = \alpha$, $\int_{w_0} p_1\, dE = 0$, $\int_{w_0} p_2\, dE = 0$, which, owing to (4.29), reduce to $\int_{w_0} p\, dE = \alpha$, $\int_{w_0} \phi_1 p\, dE = 0$, $\int_{w_0} \phi_2 p\, dE = 0$, respectively. Using these relationships, the inequality (4.31) will further simplify to

$$(4.32)\qquad \begin{aligned} &\left(\int_{w_0} \phi_1^2 p\, dE + A_{11}\alpha\right)\phi_2^2 + \left(\int_{w_0} \phi_2^2 p\, dE + A_{22}\alpha\right)\phi_1^2 \\ &\qquad - 2\left(\int_{w_0} \phi_1\phi_2 p\, dE + A_{12}\alpha\right)\phi_1\phi_2 \gtrless k_1' + k_2'\phi_1 + k_3'\phi_2. \end{aligned}$$

Here the sign $\geq$ applies in w_0 and $\leq$ outside w_0. The region described by this inequality is obviously the region outside an ellipse in the ϕ_1, ϕ_2-plane.

5. Generalization of the theory to the k-parameter case. We shall now indicate how to generalize the theory of Sections 3 and 4 to the case where we have k parameters, where $k \geq 2$. Our main task here will be to obtain a generalization of Lemma 1 and Theorem 1 of Section 4.

The power function is now designated by $\beta(\theta_1, \theta_2, \cdots, \theta_k \mid w)$, and we are testing the hypothesis that $(\theta_1, \theta_2, \cdots, \theta_k) = (0, 0, \cdots, 0)$. For brevity we write $\theta = (\theta_1, \theta_2, \cdots, \theta_k)$, so that $\beta(\theta \mid w)$ now will symbolize the power function and the hypothesis is $\theta = 0 = (0, 0, \cdots, 0)$. $\beta_i(w)$ and $\beta_{ij}(w)$ will again denote the partial derivatives of $\beta(\theta \mid w)$ evaluated at $\theta = 0$, where now i and j run from 1 to k.

We now define the generalized Gaussian curvature of $\beta(\theta \mid w)$ at $\theta = 0$ as follows:

$$K = \frac{\begin{vmatrix} \beta_{11}(w) & \cdots & \beta_{1k}(w) \\ \vdots & & \vdots \\ \beta_{k1}(w) & \cdots & \beta_{kk}(w) \end{vmatrix}}{\left(1 + \sum_{j=1}^{k} \beta_j^2(w)\right)^2} = \frac{\det B_w}{\left(1 + \sum_{j=1}^{k} \beta_j^2(w)\right)^2}, \tag{5.1}$$

where

$$B_w = \begin{pmatrix} \beta_{11}(w) & \cdots & \beta_{1k}(w) \\ \vdots & & \vdots \\ \beta_{k1}(w) & \cdots & \beta_{kk}(w) \end{pmatrix}.$$

The generalized Gaussian curvature is invariant under translation and rotation of the coordinate axes in the $(k + 1)$-dimensional space of $(\beta, \theta_1, \theta_2, \cdots, \theta_k)$. Imposing our condition of unbiasedness on $\beta(\theta \mid w)$ at $\theta = 0$ gives us $\beta_j(w) = 0$, $j = 1, \cdots, k$; and hence we have for $\beta(\theta \mid w)$ at $\theta = 0$

$$K = \det B_w. \tag{5.2}$$

In view of this discussion, our definition of a region of type D in Section 3 immediately generalizes to k parameters.

Geometrically, the region of type D may be regarded as minimizing the volume of a certain infinitesimal k-dimensional ellipsoid $\sum_{i,j=1}^{k} \beta_{ij}\theta_i\theta_j = \delta$, as explained in detail in Section 3 for the case $k = 2$.

We again assume the fundamental condition at the beginning of Section 4 is satisfied. We use the notation of (4.1), (4.2), and (4.3); i.e., $\beta_i(w) = \int_w p_i \, dE$, and $\beta_{ij}(w) = \int_w p_{ij} \, dE$, where now $i, j = 1, \cdots, k$, and we let $p(E \mid 0) = p$. In terms of these p's our conditions on a type D region are expressed by I′–IV′ of Section 4, with i and j running from 1 to k.

We thus see that in the k-parameter theory we need a theorem which tells how to characterize the structure of a region which maximizes the determinant

of a symmetric positive definite $k \times k$ matrix, whose elements are integrals over the region, subject to integral side conditions. To this end, we obtain generalizations of Lemma 1 and Theorem 1 of Section 4.

We generalize the statement of Lemma 1 by replacing 2 by k whenever a 2 occurs in the statement. Relation (4.9) is replaced by

$$\text{(5.3)} \qquad \begin{aligned} \sum_{i=1}^{k} k_{ii} g_{ii} &\geq \sum_{i=1}^{m} k_i f_i \quad \textit{in} \quad w_0, \\ \sum_{i=1}^{k} k_{ii} g_{ii} &\leq \sum_{i=1}^{m} k_i f_i \quad \textit{outside} \quad w_0, \end{aligned}$$

where

$$k_{ii} = \prod_{\substack{j=1 \\ j \neq i}}^{k} \int_{w_0} g_{jj}\, dE.$$

The proof of the lemma then proceeds exactly as it does in the case $k = 2$.

The corollary to Lemma 1 is now given in the following form:

COROLLARY. *Consider a symmetric matrix*

$$G = \begin{pmatrix} g_{11} & \cdots & g_{1k} \\ \vdots & & \vdots \\ g_{k1} & \cdots & g_{kk} \end{pmatrix}$$

whose elements are given integrable functions defined in an n-dimensional space. For any region w in this space, let

$$G_w = \begin{vmatrix} \int_w g_{11}\, dE & \cdots & \int_w g_{1k}\, dE \\ \vdots & & \vdots \\ \int_w g_{k1}\, dE & \cdots & \int_w g_{kk}\, dE \end{vmatrix}.$$

Now if w_0 is a region that satisfies the conditions (4.7), (4.8), *and* (5.3) *of the lemma, and if furthermore $\int_{w_0} g_{ij}\, dE = 0$ when $i \neq j$, then det $G_{w_0} \geq$ det G_w, where w is any region in the space for which G_w is positive definite and the conditions* (4.7) *are satisfied.*

PROOF.

$$\text{(5.4)} \qquad \det G_{w_0} = \prod_{j=1}^{k} \int_{w_0} g_{jj}\, dE \geq \prod_{j=1}^{k} \int_{w} g_{jj}\, dE \geq \det G_w,$$

where the first inequality follows from the lemma and the second is a well known inequality for positive definite matrices (see Cramér [4], p. 116).

Proceeding to Theorem 1, we generalize the statement by once again replacing 2 by k whenever a 2 occurs in the statement. Relation (4.19) is

replaced by

$$
(5.5) \qquad \begin{aligned}
\sum_{i,j=1}^{k} k_{ij} g_{ij} &\geq \sum_{i=1}^{m} k_i f_i \quad \text{in} \quad w_0, \\
\sum_{i,j=1}^{k} k_{ij} g_{ij} &\leq \sum_{i=1}^{m} k_i f_i \quad \text{outside} \quad w_0,
\end{aligned}
$$

where k_{ij} is the (i, j) element in the adjoint matrix of G_{w_0}. We note that $\sum_{i,j=1}^{k} k_{ij}g_{ij} = \text{trace } [(\text{adj } G_{w_0})G]$, where adj G_{w_0} denotes the adjoint matrix of G_{w_0}. The proof of the theorem then proceeds exactly as it does in the case $k = 2$.

Regions of type D in the k-parameter theory remain invariant under transformations of the parameter space which are locally one-to-one and twice differentiable with continuous partial derivatives. This result is obtained by a direct and immediate generalization of Theorem 2 in Section 4.

As in the two-parameter theory, the inequalities which must hold within and outside the unbiased critical region of type D can frequently be simplified if we express them in terms of the derivatives of $\log p(E \mid \theta)$. We write:

$$
(5.6) \qquad \begin{aligned}
\phi_t &= \left.\frac{\partial \log p(E \mid \theta)}{\partial \theta_t}\right|_{\theta=0}, \\
\phi_{ts} &= \left.\frac{\partial^2 \log p(E \mid \theta)}{\partial \theta_t \partial \theta_s}\right|_{\theta=0},
\end{aligned}
$$

where $t, s = 1, 2, \cdots, k$. In particular the simplification will be considerable if

$$
(5.7) \qquad \phi_{ts} = A_{ts} + \sum_{j=1}^{k} B_{tsj}\phi_j, \qquad t, s = 1, 2, \cdots, k,
$$

where A_{ts} and the B_{tsj} are independent of the sample point E but may depend on θ. An unbiased critical region of type D found by application of Theorem 1 will then be the outside of an ellipsoid in the space of the ϕ_t, $t = 1, \cdots, k$.

6. Examples. Suppose that the joint density functions specified by the admissible hypotheses are all of the form

$$
(6.1) \qquad \begin{aligned}
&p(E \mid \mu_1, \mu_2) \\
&= \frac{1}{(2\pi)^{\frac{1}{2}(n_1+n_2)}\sigma_1^{n_1}\sigma_2^{n_2}} \exp\left[-\frac{1}{2}\left\{\frac{1}{\sigma_1^2}\sum_{i=1}^{n_1}(x_i - \mu_1)^2 + \frac{1}{\sigma_2^2}\sum_{i=n_1+1}^{n_1+n_2}(x_i - \mu_2)^2\right\}\right],
\end{aligned}
$$

with known σ_1 and σ_2, for $-\infty < x_i < \infty$, $i = 1, 2, \cdots, (n_1 + n_2)$. Thus it is assumed that the observations represent two samples of n_1 and n_2 individuals respectively, randomly and independently drawn from two normal populations with *known* standard deviations σ_1 and σ_2 respectively and with *unknown* means equal respectively to μ_1 and μ_2. The simple hypothesis H_0 to be tested is that $(\mu_1, \mu_2) = (0, 0)$. We shall find the unbiased critical region of type D for testing H_0.

The joint density function (6.1), as is well known, satisfies all the conditions required in the present theory. Making some simple calculations and substituting into (4.32), we obtain

$$
\begin{aligned}
f(\bar{x}_1, \bar{x}_2) = &\left(\int_{w_0}\left(\frac{n_1\bar{x}_1}{\sigma_1^2}\right)^2 p\,dE - \frac{n_1\alpha}{\sigma_1^2}\right)\left(\frac{n_2\bar{x}_2}{\sigma_2^2}\right)^2 \\
&+ \left(\int_{w_0}\left(\frac{n_2\bar{x}_2}{\sigma_2^2}\right)^2 p\,dE - \frac{n_2\alpha}{\sigma_2^2}\right)\left(\frac{n_1\bar{x}_1}{\sigma_1^2}\right)^2 \\
&- 2\left(\int_{w_0}\frac{n_1n_2\bar{x}_1\bar{x}_2}{\sigma_1^2\sigma_2^2}\,p\,dE\right)\left(\frac{n_1n_2\bar{x}_1\bar{x}_2}{\sigma_1^2\sigma_2^2}\right) - k_1' - k_2'\frac{n_1\bar{x}_1}{\sigma_1^2} - k_3'\frac{n_2\bar{x}_2}{\sigma_2^2} \geq 0
\end{aligned}
\tag{6.2}
$$

as the inequality defining the critical region w_0 we seek, providing it can be found by our methods, where $\bar{x}_1$ and $\bar{x}_2$ denote the means of the two samples. It is seen that w_0 is bounded by a surface corresponding to the equation $f(\bar{x}_1, \bar{x}_2) = \text{constant}$, and that, therefore, the conditions $\int_{w_0} p\,dE = \alpha$, $\int_{w_0} \phi_1 p\,dE = 0$, $\int_{w_0} \phi_2 p\,dE = 0$, which the critical region has to satisfy, and also the integrals involved in (4.32) can be expressed by means of integrals taken over a region w_0' in the plane of $\bar{x}_1$ and $\bar{x}_2$, determined by the same inequality (6.2). Of course, instead of the original joint density function $p(E \mid 0, 0)$, we shall have that of $\bar{x}_1$ and $\bar{x}_2$. We further simplify our notation by introducing, instead of $\bar{x}_1$ and $\bar{x}_2$, the variables

$$
u = \sqrt{n_1}\,\bar{x}_1/\sigma_1, \qquad v = \sqrt{n_2}\,\bar{x}_2/\sigma_2. \tag{6.3}
$$

Our problem will now be to guess a region w_0'' in the u, v-plane and then see if we can determine the constants k_i' so that the plane region determined by the inequality

$$
\begin{aligned}
\frac{n_1n_2}{\sigma_1^2\sigma_2^2}\Bigg[&\left(\iint_{w_0''} u^2p(u, v)\,du\,dv - \alpha\right)v^2 + \left(\iint_{w_0''} v^2p(u, v)\,du\,dv - \alpha\right)u^2 \\
&- 2\left(\iint_{w_0''} uvp(u, v)\,du\,dv\right)uv\Bigg] \geqq k_1' + k_2'\frac{\sqrt{n_1}}{\sigma_1}u + k_3'\frac{\sqrt{n_2}}{\sigma_2}v
\end{aligned}
\tag{6.4}
$$

will be the region w_0'', where w_0'' satisfies the following conditions:

$$
\iint_{w_0''} p(u, v)\,du\,dv = \alpha; \tag{6.5}
$$

$$
\iint_{w_0''} up(u, v)\,du\,dv = 0, \qquad \iint_{w_0''} vp(u, v)\,du\,dv = 0; \tag{6.6}
$$

$$
\begin{pmatrix}
\iint_{w_0''} u^2p(u, v)\,du\,dv - \alpha & \iint_{w_0''} uvp(u, v)\,du\,dv \\
\iint_{w_0''} uvp(u, v)\,du\,dv & \iint_{w_0''} v^2p(u, v)\,du\,dv - \alpha
\end{pmatrix}
\tag{6.7}
$$

is positive definite, where

$$(6.8)\qquad p(u, v) = (2\pi)^{-1} \exp\left[-\tfrac{1}{2}(u^2 + v^2)\right].$$

(6.5) is the condition of size; and (6.6), (6.7) are the conditions of unbiasedness. If we have such a region, then by Theorem 1, w_0'' is an unbiased critical region of type D for testing $(\mu_1, \mu_2) = (0, 0)$. In the u, v-plane, the likelihood ratio test indicates the region $u^2 + v^2 \geq a^2$ for testing H_0, where a^2 is determined so as to give size α to the test. Since u^2 and v^2 are each independently distributed as χ^2 with one degree of freedom, $u^2 + v^2$ is distributed as χ^2 with two degrees of freedom and so a^2 can be obtained from a χ^2-table. We shall take $u^2 + v^2 \geq a^2$ as the region w_0'' and shall verify that k_1', k_2', k_3' in (6.4) can indeed be determined so as to give rise to this region. We will also see that (6.7) is satisfied for this region. Then since $u^2 + v^2 \geq a^2$ obviously satisfies the condition (6.6) by symmetry considerations, and a^2 has been determined so as to satisfy (6.5), this will prove that $u^2 + v^2 \geq a^2$ is an unbiased critical region of type D for testing H_0.

One can easily verify that

$$(6.9)\qquad \iint_{u^2+v^2\geq a^2} u^2 p(u, v)\, du\, dv = \iint_{u^2+v^2\geq a^2} v^2 p(u, v)\, du\, dv = \alpha(1 + \tfrac{1}{2}a^2);$$

and since $p(u, v)$ is an even function,

$$(6.10)\qquad \iint_{u^2+v^2\geq a^2} uvp(u, v)\, du\, dv = 0.$$

In view of these relations, we see that the matrix in (6.7) is

$$\begin{pmatrix} \alpha a^2/2 & 0 \\ 0 & \alpha a^2/2 \end{pmatrix},$$

which is obviously positive definite. Also, (6.4) can now be written as

$$(6.11)\qquad \frac{n_1 n_2}{\sigma_1^2 \sigma_2^2}\left[\frac{\alpha a^2}{2}(u^2 + v^2)\right] \geq k_1' + k_2' \frac{\sqrt{n_1}}{\sigma_1} u + k_3' \frac{\sqrt{n_2}}{\sigma_2} v.$$

If we choose $k_1' = n_1 n_2 \alpha a^4/(2\sigma_1^2\sigma_2^2)$, $k_2' = 0$, $k_3' = 0$, the inequality (6.11) becomes

$$(6.12)\qquad u^2 + v^2 \geqq a^2,$$

and this proves our result.

This result can in turn be used to find an unbiased critical region of type D for testing a simple hypothesis about the means of a bivariate normal population with known covariance matrix, since it is possible by an orthogonal transformation of variables to transform this problem into the one we have solved.

The result of (6.12) can also be immediately extended to find an unbiased critical region of type D for testing a simple hypothesis about the means of k independent normal populations with known variances; and then this latter

result can be used to find a type D region for testing a simple hypothesis about the means of a k-variate normal distribution with known covariance matrix. The type D regions in these cases turn out to be the likelihood ratio tests.

My attempts to find an unbiased critical region of type D for testing a simple hypothesis about the mean and variance of a univariate normal distribution on the basis of a random sample of size n were unsuccessful because I was unable to evaluate the integrals occurring on the left side of our basic inequality (4.19) over the conjectured region; there were also other difficulties involved. One can, however, use the result of (6.12) for large sample sizes to approximate a type D region for testing the simple hypothesis $(\mu, \sigma^2) = (\mu_0, \sigma_0^2)$. Since $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ and $(s')^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$ are joint sufficient statistics for μ and σ^2, just as we reduced the problem of testing a simple hypothesis about the means of two normal populations to a problem in the $\bar{x}_1$, $\bar{x}_2$-plane by use of (6.2), so we can reduce the problem of testing $(\mu, \sigma^2) = (\mu_0, \sigma_0^2)$ to a problem in the $\bar{x}$, $(s')^2$-plane. The density function of $\bar{x}$ is normal with mean μ_0 and variance σ_0^2/n under the null hypothesis, and the density function of $(n-1)(s')^2/\sigma_0^2$ is that of χ^2 with $(n-1)$ degrees of freedom under the null hypothesis; since $\bar{x}$ and $(s')^2$ are independently distributed in a normal population, we can use these two density functions immediately to obtain the joint density function of $\bar{x}$ and $(s')^2$. The problem of finding a type D region in the $\bar{x}$, $(s')^2$-plane is, however, the one I was unable to solve. But we know that $(n-1)(s')^2/\sigma_0^2$ has a χ^2 distribution with mean $(n-1)$ and variance $2(n-1)$ and we also know that a χ^2 distribution with m degrees of freedom is asymptotically normal with mean m and variance $2m$; hence we know that $(s')^2$ is asymptotically normally distributed with mean σ_0^2 and variance $2\sigma_0^4/(n-1)$ under the null hypothesis. If we approximate the density function of $(s')^2$ by a normal density function with mean σ_0^2 and variance $2\sigma_0^4/(n-1)$, and let

$$u = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma_0}, \qquad v = \frac{\sqrt{n-1}((s')^2 - \sigma_0^2)}{\sqrt{2}\sigma_0^2}, \tag{6.13}$$

then with this approximation our problem becomes that of finding a region w_0' in the u, v-plane satisfying (6.4), subject to conditions (6.5)–(6.7), where $p(u, v)$ is given by (6.8), and in (6.4) $n_1 = n$, $n_2 = (n-1)$, $\sigma_1 = \sigma_0$, $\sigma_2 = \sqrt{2}\sigma_0^2$. For this problem we have seen that the solution is given by (6.12). In the $\bar{x}$, $(s')^2$-plane this gives the region

$$\frac{n(\bar{x} - \mu_0)^2}{\sigma_0^2} + \frac{(n-1)((s')^2 - \sigma_0^2)^2}{2\sigma_0^4} \geq a^2, \tag{6.14}$$

where a^2 is determined from a χ^2-table with two degrees of freedom. For large sample sizes this region should be a good approximation to an unbiased critical region of type D for testing $(\mu, \sigma^2) = (\mu_0, \sigma_0^2)$.

7. Remarks on the theory of testing composite hypotheses with two or more constraints. A composite hypothesis with k constraints is a hypothesis which

specifies the values of k parameters out of a total of s parameters, where $k < s$. At present the theory of composite hypotheses with two or more constraints is in much less satisfactory shape than the theory of composite hypotheses with one constraint. (For the latter see Scheffé [5] and Lehmann [6].) We can define an unbiased critical region of type E for testing a composite hypothesis with k constraints ($k \geq 2$, $k < s$) as follows:

DEFINITION. *Let* $\Theta = (\theta_1, \theta_2, \cdots, \theta_k, \theta_{k+1}, \cdots, \theta_s) = (\theta_1, \theta_2, \cdots, \theta_k, \tau)$ *denote the parameter point in the parameter space* Ω *which is a subset of an s-dimensional Euclidean space, where* $\tau = (\theta_{k+1}, \cdots, \theta_s)$ *denotes the nuisance parameters (i.e., the parameters unspecified by the hypothesis). The hypothesis* H_0 *states* Θ *lies in the k-dimensional subspace* ω *of* Ω *defined by* $\theta = \theta_0$, *where* $\theta = (\theta_1, \theta_2, \cdots, \theta_k)$ *and* $\theta_0 = (\theta_{10}, \theta_{20}, \cdots, \theta_{k0})$. *Then* w_0 *is said to be an unbiased critical region of type E for testing* H_0 *if for all* θ *in* ω *(i.e., all* (θ_0, τ)*):*

I. $\beta(\theta_0, \tau \mid w_0) = \alpha$, *where* α *is independent of* τ;

II. $\beta_i(\theta_0, \tau \mid w_0) = 0$ *for* $i = 1, \cdots, k$;

III. $$\begin{pmatrix} \beta_{11}(\theta_0, \tau \mid w_0) & \cdots & \beta_{1k}(\theta_0, \tau \mid w_0) \\ \vdots & & \vdots \\ \beta_{k1}(\theta_0, \tau \mid w_0) & \cdots & \beta_{kk}(\theta_0, \tau \mid w_0) \end{pmatrix} = B_{w_0} \text{ is positive definite;}$$

IV. *det* $B_{w_0} \geq$ *det* B_w *for any region* w *satisfying* I–III.

These regions of type E should prove useful in the further development of the theory of composite hypotheses with two or more constraints.

I am deeply indebted to Professor Henry Scheffé for having suggested this line of research to me and for numerous very helpful suggestions he has given me in the course of pursuing it.

REFERENCES

[1] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses, part I," *Stat. Res. Memoirs*, Vol. 1 (1936), pp. 1–37.

[2] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses, parts II and III," *Stat. Res. Memoirs*, Vol. 2 (1938), pp. 25–57.

[3] L. EISENHART, *Differential Geometry*, Ginn and Co., 1909.

[4] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[5] H. SCHEFFÉ, "On the theory of testing composite hypotheses with one constraint," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 280–293.

[6] E. L. LEHMANN, "On optimum tests of composite hypotheses with one constraint," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 473–494.

DESIGNS FOR TWO-WAY ELIMINATION OF HETEROGENEITY

By S. S. Shrikhande

University of North Carolina

1. Introduction and summary. Sometimes in a design the position within the block is important as a source of variation, and the experiment gains in efficiency by eliminating the positional effect. The classical example is due to Youden in his studies on the tobacco mosaic virus [1]. He found that the response to treatments also depends on the position of the leaf on the plant. If the number of leaves is sufficient so that every treatment can be applied to one leaf of a tree, then we get an ordinary Latin square, in which the trees are columns and the leaves belonging to the same position constitute the rows. But if the number of treatments is larger than the number of leaf positions available, then we must have incomplete columns. Youden used a design in which the columns constituted a balanced incomplete block design, whereas the rows were complete. These designs are known as Youden's squares, and can be used when two-way elimination of heterogeneity is desired.

In Fisher and Yates statistical tables [2] balanced incomplete block designs in which the number of blocks b is equal to the number of treatments v have been used to obtain Youden's squares, and the authors state that "in all cases of practical importance" it has been found possible to convert balanced incomplete blocks of the above kind to a Youden's square by so ordering the varieties in the blocks that each variety occurs once in each position. F. W. Levi noted ([3], p. 6) that this reordering can always be done, in virtue of a theorem given by Konig [4] which states that an even regular graph of degree m is the product of m regular graphs of degree 1. Smith and Hartley [5] give a practical procedure for converting balanced incomplete blocks with $b = v$ into Youden's squares.

In this paper I have considered some general classes of designs for two-way elimination of heterogeneity. In Section 3 balanced incomplete block designs for which $b = mv$ have been used to obtain two-way designs in which each treatment occurs in a given position m times. The case $m = 1$ gives Youden's squares. In Section 4 it has been shown that balanced incomplete block designs for which b is not an integral multiple of v can be used to obtain designs for two-way elimination of heterogeneity in which there are two accuracies (i.e., some pairs of treatments are compared with one accuracy, while other pairs are compared with a different accuracy) as in the case of lattice designs for one-way elimination of heterogeneity. In Sections 5 and 6 partially balanced designs have been used to obtain two-way designs with two accuracies. In every case the method of analysis and tables of actual designs have been given.

2. Notation and preliminaries. Consider a two-way design with k rows and b columns. Let there be v treatments altogether, and let n_{ij} denote the number

of times the treatment i occurs in row j, and n'_{ic} the number of times it occurs in the column c. If y_{jc} is the yield for the jth row and cth column, the mathematical model assumed will be

$$y_{jc} = g + t_i + b_j + p_c + e_{jc}, \tag{2.0}$$

where t_i is the effect of the treatment i occurring in the row j and column c, b_j and p_c are the effects of the jth row and cth column, respectively, and e_{jc} is a random variable which is distributed $N(0, \sigma^2)$ independently for each value of j and c.

Let T_i, B_j and B'_c denote respectively the totals of the yields corresponding to the treatment i, row j and column c. Put

$$Q_i = T_i - \frac{1}{b}\Sigma_j n_{ij} B_j - \frac{1}{k}\Sigma_c n'_{ic} B'_c + \frac{r_i G}{bk}, \tag{2.1}$$

where G is the grand total of all the yields and r_i is the number of replications of the ith treatment. Q_i is called the adjusted yield of the ith treatment.

Let us set

$$c_{ii} = r_i\left(1 + \frac{r_i}{bk}\right) - \frac{1}{b}\Sigma_j n_{ij}^2 - \frac{1}{k}\Sigma_c n_{ic}'^2, \tag{2.2}$$

$$c_{iu} = -\frac{1}{b}\Sigma_j n_{ij} n_{uj} - \frac{1}{k}\Sigma_c n'_{ic} n'_{uc} + \frac{r_i r_u}{bk}, \qquad i \neq u. \tag{2.3}$$

It can be easily shown that the rank of the matrix (c_{iu}) is at the most equal to $v - 1$. We shall suppose that the parameters entering in the design are such that rank (c_{iu}) is actually equal to $v - 1$. In this case the design is said to be connected. The best unbiased linear estimate of any contrast

$$l_1 t_1 + l_2 t_2 + \cdots + l_v t_v, \qquad \Sigma_i l_i = 0, \tag{2.4}$$

is obtained by solving the normal equations

$$c_{i1} t_1 + c_{i2} t_2 + \cdots + c_{iv} t_v = Q_i, \qquad i = 1, 2, \cdots, v, \tag{2.5}$$

and substituting the values in the contrast. The t's are determined up to an arbitrary constant, and may be made unique by using the constraint

$$t_1 + t_2 + \cdots + t_v = 0. \tag{2.6}$$

Let $\hat{t}_1, \hat{t}_2, \cdots, \hat{t}_v$ be any solution of (2.4). Then the analysis of variance table for the design will be Table I. Detailed proofs of the facts stated in this section can be worked out along the lines indicated by Bose [6].

3. Designs with complete rows in which every treatment occurs in a row *m* times. Consider a two-way design in which the columns form a balanced incomplete block design with parameters v, b, r, k, λ, where v is the number of treatments, b is the number of blocks, k is the block size, r is the number of replications of each treatment, and λ is the number of times any two treatments

occur together in the same column. Then

$$\Sigma_c n_{ic}'^2 = r_i = r, \qquad i = 1, 2, \cdots, v, \tag{3.0}$$

$$\Sigma_c n_{ic}' n_{uc}' = \lambda, \qquad i, u = 1, 2, \cdots, v; i \neq u. \tag{3.1}$$

Consider the matrix $N = (n_{ij})$ of v rows and k columns, n_{ij} being the number already defined. The matrix N is intrinsically associated with the positions of the treatments within the columns of the design, and depends on the parameters of the design only inasmuch as each column adds up to b and each row to

TABLE I

Analysis of variance for a two-way design

Source of variation	d.f.	Sum of squares	Mean square
Treatment contrasts eliminating rows and columns	$v - 1$	$S_t^2 = \Sigma_i \hat{t}_i Q_i$	$s_t^2 = \frac{S_t^2}{v-1}$
Row contrasts ignoring treatments	$k - 1$	$\frac{1}{b}\Sigma_i B_i^2 - \frac{G^2}{bk}$	
Column contrasts ignoring treatments	$b - 1$	$\frac{1}{k}\Sigma_c B_c'^2 - \frac{G^2}{bk}$	
Error	$(b - 1)(k - 1) - (v - 1)$	S_e^2(by subtraction)	$s_e^2 = \frac{S_e^2}{(b-1)(k-1)-(v-1)}$
Total	$bk - 1$	$\Sigma_{i,c} y_{ic}^2 - \frac{G^2}{bk}$	

$F = \frac{s_t^2}{s_e^2}$, d.f. $v - 1$, $(b - 1)(k - 1) - (v - 1)$.

r. Let b be an integral multiple of v, so that $b = mv$. Then $r = mk$. By suitable interchanges of treatments in the same column of the design, the matrix N can be so modified that

$$n_{ij} = m, \qquad i = 1, 2, \cdots, v; j = 1, 2, \cdots, k, \tag{3.2}$$

since the procedure of Smith and Hartley [5] can be easily generalized in the following manner to cover the case $m \neq 1$.

If $n_{ij} = m$ is not satisfied for all values of i, j, we define

$$\begin{aligned} m_{ij} &= m - n_{ij} && \text{if } m > n_{ij}, \\ &= 0 && \text{if } m \leq n_{ij}, \\ M &= \Sigma m_{ij}. \end{aligned}$$

Then, following the Smith-Hartley procedure, only slight modifications in the argument show that we can find an interchange or system of interchanges within

columns which would reduce M by at least unity. Successive applications of this process give the desired result, since $M = 0$ implies $n_{ij} = m$ for all i, j.

Now we have

$$\Sigma_j n_{ij}^2 = km^2, \qquad i = 1, 2, \cdots, v, \tag{3.3}$$

$$\Sigma_j n_{ij} n_{uj} = km^2, \qquad i, u = 1, 2, \cdots, v; i \neq u. \tag{3.4}$$

Under the restraint (2.6) the normal equations (2.5) become

$$\left(r - \frac{r}{k} + \frac{\lambda}{k}\right) t_i = Q_i, \tag{3.5}$$

which are exactly the same as for balanced incomplete block designs (cf. Bose [6]). Hence $\hat{t}_i$, the estimate of t_i, is Q_i/rE, where $E = u\lambda/kr$, and the analysis of variance can be obtained by substituting this value of $\hat{t}_i$ in the table at the end of Section 2. Also

$$V(\hat{t}_i - \hat{t}_u) = 2\sigma^2/rE. \tag{3.6}$$

When a cyclic or multicyclic solution of a balanced incomplete block design is available the matrix N already obeys the condition (3.2), and an actual application of the Smith-Hartley process is unnecessary. Only the designs with $r \leq 10$ are practically important, and cyclic or multicyclic solutions for all but three of these designs are available in the tables of Fisher and Yates [2], and in a paper by Rao [7]. The solution for the three missing cases are given in Table II. The solution for the design 1 of Table II is obtained by modifying the corresponding solution by Bose [8], and the solutions for designs 2 and 3 are obtained from the corresponding solutions by Bhattacharya [9], [10]. The designs considered here may be called extended Youden's square designs when $m \geq 1$.

In Table II, instead of giving the design in the row-column form, it is convenient to give the blocks corresponding to the columns. The row position is then given by the position within the block. This convention will be adopted throughout the paper. In many cases it is possible to represent the designs compactly by developing a set of blocks from one block cyclically. The fonowing convention will be adopted for this purpose. To develop the block $(a, b, \cdots, x)$ cyclically (mod g), we write down the set of g blocks $(a + t, b + t, \cdots, x + t)$, $t = 0, 1, 2, \cdots, g - 1$, and then reduce every number appearing in the blocks to lie between 1 and g (both inclusive) by subtracting g whenever a number appearing in the blocks exceeds g. In certain cases to each number between 1 and g there correspond m treatments instead of one. The treatments corresponding to c being denoted by $c_1, c_2, \cdots, c_m$. In this case in developing the blocks suffixes are left invariant. Thus by developing $(1_1, 5_2, 4_1)$ cyclically (mod 5), we get $(1_1, 5_2, 4_1)$, $(2_1, 1_2, 5_1)$, $(3_1, 2_2, 1_1)$, $(4_1, 3_2, 2_1)$, $(5_1, 4_2, 3_1)$.

Sometimes treatments are represented by compound symbols (a, b) with or without suffixes, and we have to develop a block (mod g_1, g_2). This can be done analogously. For example, by developing (mod 3, 3) the block

$$[(2, 1)_1, (1, 2)_1, (2, 2)_2, (1, 1)_2],$$

we get the nine blocks

$$[(2,1)_1, (1,2)_1, (2,2)_2, (1,1)_2], \quad [(3,1)_1, (2,2)_1, (3,2)_2, (2,1)_2],$$

$$[(1,1)_1, (3,2)_1, (1,2)_2, (3,1)_2], \quad [(2,2)_1, (1,3)_1, (2,3)_2, (1,2)_2],$$

$$[(3,2)_1, (2,3)_1, (3,3)_2, (2,2)_2], \quad [(1,2)_1, (3,3)_1, (1,3)_2, (3,2)_2],$$

$$[(2,3)_1, (1,1)_1, (2,1)_2, (1,3)_2], \quad [(3,3)_1, (2,1)_1, (3,1)_2, (2,3)_2],$$

$$[(1,3)_1, (3,1)_1, (1,1)_2, (3,3)_2].$$

TABLE II

Some extended Youden's square designs

Serial no.	Parameters: v, b, r, k, λ	Blocks
1	10, 30, 9, 3, 2	$(5_2, 1_2, 2_1), (1_1, 5_2, 4_1), (2_1, 3_1, 5_2), (1_1, 4_1, 2_2), (2_2, 3_1, 2_1), (5_2, 2_2, 5_1)$; other blocks are obtained by developing (mod 5), keeping the suffixes fixed.
2	25, 25, 9, 9, 3	(5, 1, 23, 6, 20, 12, 17, 2, 11), (18, 21, 5, 7, 10, 24, 3, 12, 1), (15, 2, 9, 10, 1, 21, 25, 17, 16), (23, 22, 11, 9, 3, 18, 1, 16, 14), (24, 13, 2, 14, 7, 8, 22, 1, 17), (8, 25, 20, 3, 6, 1, 13, 18, 15), (20, 4, 3, 17, 8, 10, 7, 23, 9), (21, 8, 24, 11, 4, 6, 2, 9, 18), (14, 12, 13, 4, 17, 25, 21, 11, 3), (3, 24, 17, 22, 15, 5, 16, 4, 6), (25, 5, 18, 20, 16, 14, 4, 7, 2), (22, 19, 1, 5, 25, 11, 10, 8, 4), (19, 14, 6, 13, 9, 17, 18, 5, 10), (1, 20, 15, 12, 19, 4, 9, 14, 24), (16, 7, 4, 1, 13, 23, 6, 21, 19), (12, 16, 8, 23, 24, 9, 5, 25, 13), (9, 3, 25, 19, 22, 2, 12, 6, 7), (7, 17, 12, 18, 11, 16, 15, 19, 8), (13, 11, 10, 16, 2, 3, 19, 24, 20), (6, 10, 16, 8, 12, 20, 14, 22, 21), (2, 23, 21, 15, 14, 19, 8, 3, 5), (10, 6, 14, 24, 23, 7, 11, 15, 25), (17, 18, 19, 25, 21, 22, 24, 20, 23), (4, 15, 22, 2, 18, 13, 23, 10, 12), (11, 9, 7, 21, 5, 15, 20, 13, 22)
3	31, 31, 10, 10, 3	(1, 2, 28, 15, 9, 11, 8, 16, 18, 4), (2, 3, 22, 16, 10, 17, 9, 19, 5, 12), (3, 4, 23, 6, 17, 13, 10, 18, 11, 20), (4, 5, 24, 18, 12, 21, 11, 14, 19, 7), (5, 6, 25, 19, 13, 8, 12, 20, 15, 1), (6, 7, 26, 20, 14, 9, 16, 21, 2, 13), (7, 1, 27, 21, 8, 10, 14, 17, 3, 15), (9, 12, 6, 1, 27, 18, 29, 26, 17, 24), (10, 13, 7, 2, 29, 25, 19, 27, 28, 18), (11, 14, 1, 3, 19, 20, 26, 28, 29, 22), (12, 8, 2, 4, 20, 29, 23, 22, 21, 27), (13, 9, 3, 5, 21, 15, 28, 23, 24, 29), (14, 10, 4, 24, 15, 16, 22, 29, 6, 25), (8, 11, 5, 29, 16, 23, 25, 7, 26, 17), (15, 24, 20, 11, 2, 27, 5, 10, 30, 26), (16, 25, 21, 12, 3, 28, 30, 11, 27, 6), (17, 26, 15, 13, 30, 22, 7, 12, 4, 28), (18, 27, 16, 14, 5, 30, 1, 13, 22, 23), (19, 28, 17, 8, 6, 14, 2, 24, 23, 30), (20, 22, 18, 9, 7, 24, 3, 30, 25, 8), (21, 23, 19, 30, 1, 26, 4, 15, 9, 10), (24, 16, 8, 27, 26, 3, 31, 4, 13, 19), (25, 17, 9, 28, 31, 4, 27, 5, 20, 14), (26, 18, 10, 22, 28, 31, 21, 6, 8, 5), (27, 19, 11, 23, 22, 6, 15, 9, 7, 31), (28, 20, 12, 31, 23, 7, 24, 1, 10, 16), (22, 21, 13, 25, 24, 1, 17, 2, 31, 11), (23, 15, 14, 26, 25, 2, 18, 31, 12, 3), (29, 30, 31, 7, 4, 5, 6, 3, 1, 2), (30, 31, 29, 10, 11, 12, 13, 8, 14, 9), (31, 29, 30, 17, 18, 19, 20, 15, 16, 21)

4. Other two-way designs obtained from balanced incomplete block designs. Balanced incomplete block designs in which the number of blocks (columns) is not an integral multiple of the number of treatments can be used to give designs with two accuracies for two-way elimination of heterogeneity. This is due to the fact that by suitable interchange of treatments in various columns it has been possible in every known case where $r \leq 10$ to express the design in a form such that in the matrix N, already referred to,

$$\Sigma_j n_{ij}^2 = \mu_2, \qquad i = 1, 2, \cdots, v, \tag{4.00}$$

$$\Sigma_j n_{ij} n_{uj} = \mu_e, \qquad i, u = 1, 2, \cdots, v; i \neq u, \tag{4.01}$$

where the treatments i and u are e-associates. These associates are similar to the associates defined by Bose and Nair [11]. Thus with respect to any treatment whatsoever, all the rest can be divided into two groups of associates with n_1 in the first group and n_2 in the second. If two treatments are e-associates, the number of treatments which are f-associates of the first and g-associates of the second is p_{fg}^e, independent of the particular pair of treatments started with. The relation of associates is reciprocal. The relations between the parameters can be derived, following Bose and Nair, as

$$\Sigma_e n_e = v - 1, \tag{4.1}$$

$$\Sigma_g p_{fg}^e = n_f \qquad \text{when } e \neq f, \tag{4.20}$$

$$= n_f - 1 \qquad \text{when } e = f, \tag{4.21}$$

$$n_e p_{fg}^e = n_f p_{ge}^f = n_g p_{ef}^g. \tag{4.3}$$

The normal equations for the estimation of treatment effects are (2.5), with

$$c_{ii} = r\left(1 - \frac{1}{k} + \frac{r}{bk}\right) - \frac{\mu_2}{b} = \alpha, \qquad i = 1, 2, \cdots, v, \tag{4.40}$$

$$c_{iu} = \frac{r}{v} - \frac{\lambda}{k} - \frac{\mu_e}{k} = \beta_e, \qquad i, u = 1, 2, \cdots, v; i \neq u, \tag{4.41}$$

where the treatments i and u are e-associates.

Following the method indicated by Bose [6], a solution of the normal equations is found to be

$$\alpha \hat{t}_i = Q_i - (\beta_1 A_{11} + \beta_2 A_{21}) Q_1(i) - (\beta_1 A_{12} + \beta_2 A_{22}) Q_2(i), \tag{4.5}$$

where $Q_e(i)$ denotes the sum of the Q's for all the e-associates of the treatment i, and (A_{ef}) is the inverse of the matrix (a_{ef}) whose elements are given by

$$a_{ef} = \alpha\delta_{ef} + \beta_e n_e + \beta_1 p_{e1}^f + \beta_2 p_{e f}^2, \qquad e, f = 1, 2, \tag{4.6}$$

where $\delta_{ef} = 1$ or 0, according as $e = f$ or $e \neq f$.

The analysis of variance can be obtained by substituting for $\hat{t}_i$ in Table I

$$V(\hat{t}_i - \hat{t}_u) = \frac{2\sigma^2}{\alpha}\{1 + \beta_1 A_{1e} + \beta_2 A_{2e}\} \tag{4.7}$$

if the treatments i and u are e-associates.

The designs considered here will be said to belong to the class Y_1. The parameters of some useful designs of this class are given in Table IIIa, and the actual designs in the Table IIIb.

The ratio of the variances of the two different kinds of comparisons is given by

$$R = \frac{1 + \beta_1 A_{11} + \beta_2 A_{21}}{1 + \beta_1 A_{12} + \beta_2 A_{22}}. \tag{4.8}$$

We shall now give a number of useful designs belonging to the class Y_1. One set of designs is obtained from the orthogonal series designs with the parameters

$$v = s^2, \qquad b = s^2 + s, \qquad r = s + 1, \qquad k = s, \qquad \lambda = 1, \tag{4.90}$$

the other parameters being

$$n_1 = s(s - 1), \qquad n_2 = s - 1, \qquad \mu_1 = s + 2, \qquad \mu_2 = s + 3, \tag{4.91}$$

$$(p_{fg}^1) = \begin{pmatrix} s(s-2) & s-1 \\ s-1 & 0 \end{pmatrix}, \qquad (p_{fg}^2) = \begin{pmatrix} s(s-1) & 0 \\ 0 & s-2 \end{pmatrix}, \tag{4.92}$$

$$R = 1 + \frac{1}{s^3 + s^2 - s}. \tag{4.93}$$

These designs are obtained by using the difference sets of Bose [12]. He has shown that if $(d_1, d_2, \cdots, d_s)$ is the difference set corresponding to s, where s is a prime or power of a prime, then a solution of the balanced incomplete block design with parameters (4.90) is obtained as follows:

(i) $s^2 - 1$ blocks are obtained by developing the block $(d_1, d_2, \cdots, d_s)$ cyclically (mod $s^2 - 1$);

(ii) $s + 1$ other blocks are obtained from the block $(0, s + 1, 2(s + 1), \cdots, (s - 2)(s + 1), \infty)$ by adding successively the numbers $1, 2, \cdots, s + 1$, where ∞ remains invariant under the addition.

To convert this solution into a two-way design of the class Y_1, we keep the $s^2 - 1$ blocks (i) unchanged. Also the first two blocks of (ii) are kept unchanged, but in the others ∞ is successively moved to the left. Finally replace ∞ by s^2. For example, the difference set corresponding to $s = 3$ is (1, 6, 7), and hence the blocks of the design corresponding to $s = 3$ are (1, 6, 7), (2, 7, 8), (3, 8, 1), (4, 1, 2), (5, 2, 3), (6, 3, 4), (7, 4, 5), (8, 5, 6); (1, 5, 9), (2, 6, 9), (3, 9, 7), (9, 4, 8).

The method of identifying the associates is easy. Divide the treatments into s groups: $(1, 2, \cdots, s)$, $(s + 1, s + 2, \cdots, 2s)$, $\cdots$, $(s^2 - s + 1, s^2 - s + 2,$

$\cdots, s^2$). Any two treatments are 1-associates if they are in the different group and 2-associates if they are in the same group.

Bose's difference sets for $s = 2, 3, 4, 5, 7, 8$, and 9 are given below.

s	*Difference set*
2	**1, 2**
3	**1, 6, 7**
4	**1, 3, 4, 12**
5	**1, 3, 16, 17, 20**
7	**1, 2, 5, 11, 31, 36, 38**
8	**1, 6, 8, 14, 38, 48, 49, 52**
9	**1, 13, 35, 48, 49, 66, 72, 74, 77**

The parameters of some other designs of the class Y_1 are given in Table IIIa. The corresponding blocks are given in Table IIIb. In each case the treatments can be divided into a number of groups such that the treatments in different groups are 1-associates, and treatments in the same group are 2-associates. These groups are also shown in Table IIIb.

TABLE IIIa

Some designs of the class Y_1: Parameters

Reference no.	v, n_1,	b, n_2,	r, μ_1,	k, μ_2,	λ, R	(p^1_{js})	(p^2_{js})
1	10, 5,	15, 4,	6, 8,	4, 10,	2, 67/65	$\begin{pmatrix} 0 & 4 \\ 4 & 0 \end{pmatrix}$	$\begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix}$
2	6, 4,	10, 1,	5, 8,	3, 9,	2, 39/38	$\begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix}$
3	8, 6,	14, 1,	7, 12,	4, 13,	3, 83/82	$\begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 6 & 0 \\ 0 & 0 \end{pmatrix}$
4	15, 10,	35, 4,	7, 16,	3, 17,	1, 171/170	$\begin{pmatrix} 5 & 4 \\ 4 & 0 \end{pmatrix}$	$\begin{pmatrix} 10 & 0 \\ 0 & 3 \end{pmatrix}$
5	10, 8,	18, 1,	9, 16,	5, 17,	4, 143/142	$\begin{pmatrix} 6 & 1 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 8 & 0 \\ 0 & 0 \end{pmatrix}$
6	16, 8,	24, 7,	9, 12,	6, 15,	3, 57/56	$\begin{pmatrix} 0 & 7 \\ 7 & 0 \end{pmatrix}$	$\begin{pmatrix} 8 & 0 \\ 0 & 6 \end{pmatrix}$

TABLE IIIb

Some designs of the class Y_1: Blocks and groups for identifying associates

Reference no.	Blocks	Groups
1	Develop the blocks $(1_1, 2_1, 4_2, 4_1)$, $(2_1, 1_2, 3_1, 4_2)$, $(1_2, 2_1, 2_2, 3_2)$, (mod 5), keeping the suffixes fixed	There are two groups. Treatments with the same suffixes belong to the same group
2	(6, 2, 3), (4, 3, 2), (3, 5, 4), (6, 5, 4), (5, 6, 2), (3, 1, 5), (2, 4, 1), (1, 6, 3), (4, 1, 6), (5, 2, 1)	There are three groups: (1, 2), (3, 4), (5, 6)
3	Develop the block (3, 5, 6, 7), (mod 7), and add the blocks (8, 2, 1, 4), (8, 5, 3, 2), (3, 8, 4, 6), (4, 8, 7, 5), (5, 6, 8, 1), (6, 7, 2, 8), (7, 1, 8, 3)	There are four groups: (1, 2), (3, 4), (5, 6), (7, 8)
4	(2, 1, 3), (4, 1, 5), (4, 6, 2), (8, 9, 1), (12, 8, 4), (8, 10, 2), (3, 13, 14), (5, 11, 14), (13, 6, 11), (14, 7, 9), (7, 11, 12), (7, 13, 10), (6, 7, 1), (2, 5, 7), (4, 7, 3), (1, 10, 11), (11, 3, 8), (2, 9, 11), (5, 8, 13), (1, 12, 13), (9, 4, 13), (12, 2, 14), (14, 6, 8), (10, 14, 4), (3, 5, 12), (15, 10, 5), (6, 15, 9), (15, 8, 7), (11, 4, 15), (13, 2, 15), (1, 14, 15), (3, 5, 6), (9, 3, 10), (5, 9, 12), (10, 12, 6)	There are three groups: (1, 2, 3, 4, 5), (6, 7, 8, 9, 10), (11, 12, 13, 14, 15)
5	(8, 2, 4, 10, 6), (7, 8, 10, 2, 1), (3, 8, 9, 4, 7), (9, 10, 1, 8, 5), (2, 5, 1, 10, 3), (10, 3, 4, 1, 6), (6, 1, 9, 5, 4), (5, 6, 8, 2, 9), (1, 6, 7, 3, 8), (4, 9, 2, 7, 10), (5, 10, 3, 9, 7), (6, 7, 2, 9, 1), (9, 1, 3, 4, 2), (4, 5, 8, 3, 2), (7, 2, 6, 5, 3), (3, 9, 10, 6, 8), (8, 7, 5, 1, 4), (10, 4, 7, 6, 5)	There are five groups: (1, 2), (3, 4), (5, 6), (7, 8), (9, 10)
6	(1, 2, 7, 8, 13, 14), (5, 13, 14, 12, 6, 11), (3, 10, 13, 9, 4, 14), (5, 6, 3, 15, 16, 4), (7, 9, 8, 10, 16, 15), (1, 2, 15, 16, 12, 11), (6, 3, 1, 15, 8, 13), (7, 15, 5, 13, 10, 12), (9, 11, 4, 13, 15, 2), (14, 7, 4, 16, 2, 5), (8, 14, 6, 9, 11, 16), (12, 3, 1, 10, 14, 16), (16, 4, 6, 1, 13, 7), (10, 8, 16, 11, 5, 13), (13, 16, 2, 12, 9, 3), (8, 5, 2, 14, 3, 15), (15, 12, 7, 6, 14, 9), (11, 4, 10, 14, 15, 1), (6, 5, 9, 2, 1, 10), (3, 1, 5, 7, 11, 9), (4, 1, 8, 5, 9, 12), (4, 7, 3, 11, 12, 8), (2, 8, 12, 4, 10, 6), (2, 6, 11, 3, 7, 10)	There are two groups: (1, 2, 3, 4, 5, 6, 7, 8), (9, 10, 11, 12, 13, 14, 15, 16)

5. Partial and extended partial Youden squares. We have seen how balanced incomplete block designs can be used for obtaining designs for two-way elimination of heterogeneity. In this and the following section we shall consider the use of partially balanced designs [11], [13] for the same purpose. The case when $b = v$ has already been considered by Bose and Kishen [14]. They call these

designs partial Youden's squares. In this section we shall consider the case $b = mv$, $r = mk$, when $m \neq 1$; we may call these designs extended partial Youden's squares.

TABLE IV

Cyclic solutions to partially balanced incomplete block designs leading to partial and extended partial Youden's squares

Reference no.	v, n_1,	b, n_2,	r, λ_1,	k, λ_2	(p^1_{fg})	(p^2_{fg})	Solution
1	13, 6,	13, 6,	3, 1,	3, 0	$\begin{pmatrix} 2 & 3 \\ 3 & 3 \end{pmatrix}$	$\begin{pmatrix} 3 & 3 \\ 3 & 2 \end{pmatrix}$	Develop (mod 13) the block (1, 3, 9).
2	15, 12,	30, 2,	6, 1,	3, 0	$\begin{pmatrix} 9 & 2 \\ 2 & 0 \end{pmatrix}$	$\begin{pmatrix} 12 & 0 \\ 0 & 1 \end{pmatrix}$	Develop (mod 15) the blocks (1, 7, 9) and (1, 12, 15).
3	15, 12,	15, 2,	4, 1,	4, 0	$\begin{pmatrix} 9 & 2 \\ 2 & 0 \end{pmatrix}$	$\begin{pmatrix} 12 & 0 \\ 0 & 1 \end{pmatrix}$	Develop (mod 15) the block (1, 3, 4, 12).
4	17, 8,	17, 8,	8, 4,	8, 3	$\begin{pmatrix} 3 & 4 \\ 4 & 4 \end{pmatrix}$	$\begin{pmatrix} 4 & 4 \\ 4 & 3 \end{pmatrix}$	Develop (mod 17) the block (1, 9, 13, 15, 16, 8, 4, 2).
5	24, 20,	24, 3,	5, 1,	5, 0	$\begin{pmatrix} 16 & 3 \\ 3 & 0 \end{pmatrix}$	$\begin{pmatrix} 20 & 0 \\ 0 & 2 \end{pmatrix}$	Develop (mod 24) the block (1, 3, 16, 17, 20).
6	25, 12,	50, 12,	6, 1,	3, 0	$\begin{pmatrix} 5 & 6 \\ 6 & 6 \end{pmatrix}$	$\begin{pmatrix} 6 & 6 \\ 6 & 5 \end{pmatrix}$	Develop (mod 5, 5) the blocks (1, 5), (1, 4), (3, 1)] and [(3, 5), (3, 2), (4, 3)].
7	26, 24,	26, 1,	9, 3,	9, 0	$\begin{pmatrix} 22 & 1 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 24 & 0 \\ 0 & 0 \end{pmatrix}$	Develop (mod 26) the block (0, 1, 2, 8, 11, 18, 20, 22, 23).
8	29, 14,	29, 14,	7, 2,	7, 1	$\begin{pmatrix} 6 & 7 \\ 7 & 7 \end{pmatrix}$	$\begin{pmatrix} 7 & 7 \\ 7 & 6 \end{pmatrix}$	Develop (mod 29) the block (1, 16, 24, 7, 25, 23, 20).
9	48, 42,	48, 5,	7, 1,	7, 0	$\begin{pmatrix} 36 & 5 \\ 5 & 0 \end{pmatrix}$	$\begin{pmatrix} 42 & 0 \\ 0 & 4 \end{pmatrix}$	Develop (mod 48) the block (1, 2, 5, 11, 31, 36, 38).
10	63, 56,	63. 6,	8, 1,	8, 0	$\begin{pmatrix} 49 & 6 \\ 6 & 0 \end{pmatrix}$	$\begin{pmatrix} 56 & 0 \\ 0 & 5 \end{pmatrix}$	Develop (mod 63) the block (1, 6, 8, 14, 38, 48, 49, 52).
11	80, 72,	80, 7,	9, 1,	9, 0	$\begin{pmatrix} 64 & 7 \\ 7 & 0 \end{pmatrix}$	$\begin{pmatrix} 72 & 0 \\ 0 & 6 \end{pmatrix}$	Develop (mod 80) the block (1, 13, 35, 48, 49, 66, 72, 74, 77).

Suppose there exists a partially balanced design with l different kinds of associates, and parameters $v, b, r, k; \lambda_1, \lambda_2, \cdots, \lambda_l; n_1, n_2, \cdots, n_l; p^e_{fg}$ $(e, f, g = 1, 2, \cdots, l)$. When $b = mv$, $r = mk$, then Smith and Hartley's process

can be used just as in Section 3 to so modify the design that each treatment occurs just m times in each row (the columns constituting the blocks). In this case we have

$$c_{ii} = r\left(1 + \frac{r}{bk}\right) - \frac{km^2}{b} - \frac{r}{k} = r\left(1 - \frac{1}{k}\right) = \alpha, \tag{5.0}$$

$$c_{iu} = -\frac{km^2}{b} - \frac{\lambda_e}{k} + \frac{r^2}{bk} = -\frac{\lambda_e}{k} = \beta_e, \tag{5.1}$$

so that the normal equations take exactly the same form as for partially balanced incomplete block designs. Hence a solution of the normal equations is given by equations (4.5) and (4.6) of Section 4, with α and β_e now given by (5.0) and (5.1). The equation (4.7) is also valid. In case there are only two kinds of associates, the ratio of the variances of the two kinds of comparisons is given by (4.8).

When a cyclic solution to a partially balanced incomplete block design with $b = mv$, $r = mk$ is available, then it can be directly used as a two-way design without further modification. A number of cyclic solutions have been given by Bose and Nair [11]. Cyclic solutions to a number of new designs are given in Table IV. In each case $l = 2$.

6. Other two-way designs obtained from partially balanced incomplete block designs. Under certain conditions it is possible to use a partially balanced design with two types of associates to give a two-way design with two types of accuracies even when b is not an integral multiple of v. The necessary condition is

$$\frac{v}{k} = \frac{b}{r} = n_1 + 1 \quad \text{or} \quad n_2 + 1. \tag{6.0}$$

In this case it has been found that in all cases of practical interest we can, by suitable interchanges within columns, arrange that

$$\Sigma_j n_{ij}^2 = d, \qquad i = 1, 2, \cdots, v, \tag{6.1}$$

$$\Sigma n_{ij} n_{uj} = \mu_e, \qquad i, u = 1, 2, \cdots, v;\ i \neq u;\ e = 1, 2, \tag{6.2}$$

where two treatments which are e-associates for the columns are also e-associates for the rows.

In this case

$$c_{ii} = r\left(1 + \frac{r}{bk}\right) - \frac{d}{b} - \frac{r}{k} = \alpha, \qquad i = 1, 2, \cdots, v, \tag{6.3}$$

$$c_{iu} = -\frac{\mu_e}{b} - \frac{\lambda_e}{k} + \frac{r^2}{bk} = \beta_e, \qquad i, u = 1, 2, \cdots, v;\ i \neq u. \tag{6.4}$$

The analysis is the same as in Section 4, the equations (4.5), (4.6), (4.7), (4.8) remaining valid but α and β now given by (6.3) and (6.4). The analysis of variance is obtained by substituting for $\hat{t}_i$ in Table I.

The designs considered here may be said to belong to class Y_2. Some designs of this class are given below in Tables Va and Vb. The parameters are given in Table Va whereas the actual solutions appear in Table Vb. In this case the representation is such that two treatments in the same group are 1-associates, whereas two treatments in different groups are 2-associates. These groups are also shown in Table Vb.

TABLE Va

Some designs of the class Y_2: Parameters

Reference no.	v, λ_1,	b, λ_2,	r, μ_1,	k, μ_2,	n_1, d	n_2	(p^1_{jk})	(p^2_{jk})
1	12, 5,	10, 2,	5, 5,	6, 4,	1, 5	10	$\begin{pmatrix} 0 & 0 \\ 0 & 10 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 \\ 1 & 8 \end{pmatrix}$
2	15, 0,	25, 1,	5, 11,	3, 7,	4, 11	10	$\begin{pmatrix} 3 & 0 \\ 0 & 10 \end{pmatrix}$	$\begin{pmatrix} 0 & 4 \\ 4 & 5 \end{pmatrix}$

TABLE Vb

Some designs of the class Y_2: Blocks and groups for identifying associates

Reference no.	Blocks	Groups
1	(8, 11, 5, 7, 1, 2), (2, 3, 1, 8, 7, 9), (3, 4, 10, 9, 2, 8), (4, 10, 11, 5, 9, 3), (5, 1, 4, 11, 10, 7), (6, 5, 8, 2, 12, 11), (9, 12, 7, 6, 3, 1), (10, 6, 2, 12, 8, 4), (11, 9, 12, 3, 6, 5), (12, 7, 6, 1, 4, 10)	There are six groups: (1, 7), (2, 8), (4, 9), (9, 10), (5, 11), (6, 12).
2	(10, 6, 4), (3, 7, 5), (11, 2, 13), (1, 9, 8), (14, 12, 15), (11, 12, 5), (3, 6, 8), (14, 7, 4), (10, 9, 13), (1, 2, 15), (1, 7, 13), (3, 2, 4), (10, 12, 8), (14, 9, 5), (11, 6, 15), (15, 3, 9), (4, 1, 12), (5, 10, 2), (8, 11, 7), (13, 14, 6), (2, 8, 14), (6, 5, 1), (9, 4, 11), (7, 5, 10), (12, 13, 3)	There are three groups: (1, 3, 10, 11, 14), (2, 6, 7, 9, 12), (4, 5, 8, 13, 15).

My sincerest thanks are due to Professor R. C. Bose, under whose guidance this research was carried out.

REFERENCES

[1] W. J. YOUDEN, "Use of incomplete block replications in estimating tobacco mosaic virus," *Contributions from Boyce Thompson Institute*, Vol. 9 (1937), pp. 317–326.
[2] R.A. FISHER AND F. YATES, *Statistical Tables*, 3rd ed., Hafner Publishing Co., New York, 1948.
[3] F. W. LEVI, *Finite Geometrical Systems*, University of Calcutta, 1942.
[4] J. KONIG, *Theorie der endlichen und unendlichen Graphen*, 1936.

[5] C. A. B. SMITH AND H. O. HARTLEY, "The construction of Youden squares," *Jour. Roy. Stat. Soc., Ser. B*, Vol. 10 (1948), pp. 262–263.
[6] R. C. BOSE, "Least square aspects of analysis of variance," mimeographed notes, Institute of Statistics, University of North Carolina.
[7] C. R. RAO, "Difference sets and combinatorial arrangements derivable from finite geometries," *Proc. Nat. Inst. Sci. India*, Vol. 12 (1946), pp. 123–135.
[8] R. C. BOSE, "On the construction of balanced incomplete block designs," *Annals of Eugenics*, Vol. 9 (1939), pp. 353–399.
[9] K. N. BHATTACHARYA, "On a new symmetrical balanced incomplete block design," *Bull. Calcutta Math. Soc.*, Vol. 36 (1944), pp. 91–96.
[10] K. N. BHATTACHARYA, "A new solution in balanced incomplete block designs," *Sankhyā*, Vol. 7 (1946), pp. 423–424.
[11] R. C. BOSE AND K. R. NAIR, "Partially balanced incomplete block designs," *Sankhyā*, Vol. 4 (1939), pp. 337–373.
[12] R. C. BOSE, "An affine analogue of Singer's theorem," *Jour. Indian Math. Soc.* (new series), Vol. 6 (1942), pp. 1–16.
[13] R. C. BOSE, "Recent work on incomplete block design in India," *Biometrics*, Vol. 3 (1947), pp. 176–178.
[14] R. C. BOSE AND K. KISHEN, "On partially balanced Youden's squares," *Science and Culture*, Vol. 5 (1939), pp. 136–137.

GENERALIZED HIT PROBABILITIES WITH A GAUSSIAN TARGET[1]

By D. A. S. Fraser

University of Toronto

1. Summary. A general discrete distribution is obtained whose random variable is the number of "hits" on a target. The target is k-dimensional and Gaussian diffuse, that is, the probability of a hit is given to within a constant factor by a Gaussian probability density function of the position of the "trajectory" in k dimensions. For a series of n rounds, the n positions of the trajectory have a multivariate Gaussian distribution. An expression is given, using Theorems 1 and 2 or 1 and 3, for the probability of r hits as a linear combination of probabilities of all hits on each possible set of rounds. Theorems 4, 5, and 6, with Theorem 1, give three limiting distributions as n, the number of rounds, tends to infinity. Theorems 7, 8, and 9, with Theorem 1, present three other limiting cases, and Theorems 10 and 1 give a time average result.

2. The problem. In [1], L. B. C. Cunningham and H. R. B. Hynd proposed a problem in multivariate statistics: to find the probability of at least one hit when an automatic gun is used against a moving target. Because of inability in aiming, the point of aim, by which we mean the centre of the distribution of the shell trajectory, will not always be the centre of the target. In fact, while the gun is being fired, the point of aim is found to wander back and forth across the target. The main complication in the problem arises in taking account of the dependence between the successive points of aim at the instants of firing.

In [1] the problem is given an approximate solution covering a partial range of parameter values and assuming the target has a circular outline.

Here the problem is modified by using a Gaussian diffuse target, a target for which the probability of a hit is given to within a constant factor by a Gaussian probability density function of the position of the trajectory. From a target which is essentially two-dimensional for aiming, the problem is generalized to a target in k dimensions, having in mind the possibility of application to other problems.

If we assume the target to be a Gaussian diffuse target and the position of the trajectory to be distributed according to a two-dimensional Gaussian distribution about the point of aim, then the probability of a "hit" as a function of the point of aim also has the form of a Gaussian diffuse target; that is, it is a constant times a Gaussian pdf of the point of aim. This will be discussed in a later paper, where the general theory will be applied to the two-dimensional problem as proposed by Cunningham and Hynd and a method of numerical evaluation considered and applied to an example.

[1] Prepared in connection with research sponsored by the Office of Naval Research.

For the general theory we shall start with the Gaussian diffuse target in terms of the point of aim, consider it in k dimensions, and call it a "success function." The point of aim is a random variable in k dimensions and will be called a prediction. If the prediction yields a hit we shall speak of a successful prediction.

The abstracted statistical problem may be stated as follows. In a series of n predictions having a joint distribution, find the probability distribution of the number R of successful predictions. Let the ith prediction be $\bar{X}_i = (X_{1i}, X_{2i}, \cdots, X_{ki}) = \{X_{\mu i}\}$ where μ ranges over the set $(1, 2, \cdots, k)$. A prediction $\bar{X}_i = \bar{x}_i$ becomes a successful prediction with probability given by the success function $s_i(\bar{x}_i)$, that is,

$$Pr\{\text{Successful prediction} \mid \bar{X}_i = \bar{x}_i\} = s_i(\bar{x}_i),$$

where $0 \leq s_i(\bar{x}_i) \leq 1$.

In the following theory the problem is solved when the predictions have a Gaussian distribution and the success functions have the form of a Gaussian diffuse target.

In the original problem of Cunningham and Hynd it was found that the horizontal and vertical components of the point of aim were reasonably independent. Consequently we shall assume independence between the values of the μth coordinate for the n predictions and values of the νth coordinate, where $\mu \neq \nu = 1, 2, \cdots, k$. The generalization omitting this independence provides little additional complication.

For each value of μ, let the set $\{X_{\mu i}\}$ have a Gaussian distribution with means $\{m_{\mu i}\}$ and covariance matrix $\| \sigma_{ij}^{(\mu)} \|$ which is positive definite. Because we shall want the probability density function for any subset of the predictions, we introduce the following notation for the inverse of the covariance matrix corresponding to a subset. For a typical subset $(i_1, i_2, \cdots, i_r)$ of the integers $(1, 2, \cdots, n)$ we shall use the symbol β_r. Then if p, q range over this subset, we have

$$\| \sigma_{pq}^{(\mu)} \|^{-1} = \| \sigma_{\beta_r(\mu)}^{pq} \| \tag{2.1}$$

as the inverse of the covariance matrix for the μth coordinates of the subset β_r of the predictions. Therefore they will have probability density element

$$\frac{| \sigma_{\beta_r(\mu)}^{pq} |^{\frac{1}{2}}}{(2\pi)^{r/2}} \exp\,[-\tfrac{1}{2}\{ \sum_{p,q\epsilon\beta_r} \sigma_{\beta_r(\mu)}^{pq}(x_{\mu p} - m_{\mu p})(x_{\mu q} - m_{\mu q})\}] \prod_{p\epsilon\beta_r} dx_{\mu p}. \tag{2.2}$$

Let the success function of the ith prediction have the following form:

$$s_i(\bar{x}_i) = C_i \exp\,[-\tfrac{1}{2} \sum_{\mu,\nu} \tau_{(i)}^{\mu\nu} x_{\mu i} x_{\nu i}], \tag{2.3}$$

where $0 \leq C_i \leq 1$, $\| \tau_{(i)}^{\mu\nu} \|$ is positive definite, and μ, ν range over the set $\{1, 2, \cdots, k\}$. There is no essential restriction in assuming that the success function is centered at the origin, since a change of origin in each k-dimensional space to center the success functions would only adjust the values $m_{\mu i}$.

3. Probabilities from expectations. To describe the distribution of R we need the probabilities of 0, 1, 2, $\cdots$, n successful predictions, that is, $Pr\{R = r\}$ for $r = 0, 1, \cdots, n$. These will not be given in the main theorems, but rather an expression for E_r defined below from which the probabilities can be calculated by well known formulas which are given in Theorem 1.

$$E_r = \sum_{i_1 < \cdots < i_r} E_{i_1 i_2 \cdots i_r} = \sum_{\beta_r} E_{\beta_r}, \tag{3.1}$$

where the summation is over all sets of r integers chosen from the set $(1, 2, \cdots, n)$. $E_{i_1 i_2 \ldots i_r}$ is the probability that predictions $i_1, i_2, \cdots, i_r$ will be successful predictions. E_r can be interpreted as the expected number of sets of r successful predictions, counted with overlapping, in our series of n predictions. This is easily seen since E_r is the sum of the probabilities for all the possible sets of r predictions.

THEOREM 1. *If E_r is defined by equation* (3.1) *and following, then the probability of* 0 *successes is*

$$Pr\{R = 0\} = 1 - E_1 + E_2 - \cdots + (-1)^n E_n, \tag{3.2}$$

and the probability of r successes is

$$\begin{aligned} Pr\{R = r\} &= \frac{1}{r!}\left\{r!E_r - \frac{(r+1)!}{1!} E_{r+1} + \cdots + (-1)^s \frac{(r+s)!}{s!} E_{r+s} \right. \\ &\qquad \left. + \cdots + (-1)^{n-r} \frac{n!}{(n-r)!} E_n \right\} \\ &= E_r - \binom{r+1}{r} E_{r+1} + \cdots + (-1)^s \binom{r+s}{r} E_{r+s} \\ &\qquad + \cdots + (-1)^{n-r} \binom{n}{r} E_n . \end{aligned} \tag{3.3}$$

These are well known formulas of probability theory.

4. The main theorem.

THEOREM 2. *Given that the success functions are Gaussian diffuse as given by* (2.3), *and that the n predictions have a Gaussian joint distribution as given by* (2.1) *and* (2.2), *then*

$$E_{\beta_r} = (\prod_p C_p) \mid \delta_{pq}\delta_{\mu\nu} + \sigma_{pq}^{(\mu)} \tau_{(p)}^{\mu\nu} \mid^{-\frac{1}{2}} \tag{4.1}$$

when all the $m_{\mu i} = 0$, and a more general formula is given by (4.4) *and* (4.5) *below.*

PROOF. Consider the following expression for E_{β_r}:

$$
\begin{aligned}
E_{\beta_r} &= Pr\{\text{predictions } i_1, i_2, \cdots, i_r \text{ will be successful predictions}\} \\
&= E\{\prod_p [C_p \exp(-\tfrac{1}{2} \sum_{\mu,\nu} \tau^{\mu\nu}_{(p)} x_{\mu p} x_{\nu p})]\} \\
&= (\prod_p C_p) \frac{\prod_\mu | \sigma^{pq}_{\beta_r(\mu)}|^{\frac{1}{2}}}{(2\pi)^{kr/2}} \\
&\quad \cdot \int \exp[-\tfrac{1}{2} \sum_{\mu,\nu,p,q} \{\tau^{\mu\nu}_{(p)} \delta_{pq} x_{\mu p} x_{\nu q} + \sigma^{pq}_{\beta_r(\mu)} \delta_{\mu\nu} (x_{\mu p} - m_{\mu p})(x_{\nu q} - m_{\nu q})\}] \prod_{\mu,p} dx_{\mu p} \\
&= (\prod_p C_p) \frac{| A |^{\frac{1}{2}}}{(2\pi)^{kr/2}} \int \exp[-\tfrac{1}{2}\{y'Ty + (y-m)'A(y-m)\}]\, dy.
\end{aligned}
$$

The matrices in the last expression are defined by

$$A = \| \sigma^{pq}_{\beta_r(\mu)} \delta_{\mu\nu} \|, \tag{4.2}$$

$$T = \| \tau^{\mu\nu}_{(p)} \delta_{pq} \|, \tag{4.3}$$

$$y = \| x_{\mu p} \|,$$

$$m = \| m_{\mu p} \|.$$

The matrices are kr by kr or kr by 1, with μ, p indicating rows and ν, q indicating columns.

$$
\begin{aligned}
E_{\beta_r} &= (\prod_p C_p) \frac{| A |^{\frac{1}{2}}}{(2\pi)^{kr/2}} \\
&\quad \cdot \int \exp[-\tfrac{1}{2}\{y'(T+A)y - y'Am - m'Ay + m'A(T+A)^{-1}Am\}]\, dy \\
&\qquad \cdot \exp[-\tfrac{1}{2}\{m'Am - m'A(T+A)^{-1}Am\}].
\end{aligned}
$$

We have completed the quadratic form by removing an appropriate factor from under the sign of integration. Integrating over the whole space, we find

$$
\begin{aligned}
E_{\beta_r} &= (\prod_p C_p) \frac{| A |^{\frac{1}{2}}}{(A+T)^{\frac{1}{2}}} \exp[-\tfrac{1}{2}(m'Am - m'A(A+T)^{-1} Am)] \\
&= (\prod_p C_p) \, | I + A^{-1}T |^{-\frac{1}{2}} \exp[-\tfrac{1}{2} m'Bm] \qquad (4.4) \\
&= (\prod_p C_p) \, | \delta_{pq}\delta_{\mu\nu} + \sigma^{(\mu)}_{pq} \tau^{\mu\nu}_{(p)} |^{-\frac{1}{2}} \exp[-\tfrac{1}{2} \sum_{\mu,\nu,p,q} m_{\mu p} B_{\mu,p,\nu,q} m_{\nu q}],
\end{aligned}
$$

where

$$
\begin{aligned}
B &= \| B_{\mu,p,\nu,q} \| \\
&= A - A(A+T)^{-1}A \\
&= A[I - (I + A^{-1}T)^{-1}] \qquad (4.5) \\
&= \| \sigma^{(\mu)}_{pq} \delta_{\mu\nu} \| \cdot [\, \| \delta_{pq}\delta_{\mu\nu} \| - \| \delta_{pq}\delta_{\mu\nu} + \sigma^{(\mu)}_{pq} \tau^{\mu\nu}_{(p)} \|^{-1}].
\end{aligned}
$$

5. Simplifications. There are two important cases given by Theorem 3 and its corollary in which we obtain a simplification in the formula for E_{β_r}.

THEOREM 3. *Given the conditions of Theorem 2 and the condition that* $\|\tau_{(i)}^{\mu\nu}\|$ *is diagonal for each* i*, the following expression is obtained for* E_{β_r}:

$$E_{\beta_r} = \left(\prod_p C_p\right) \prod_\mu | \delta_{pq} + \sigma_{pq}^{(\mu)} \tau_{(p)}^{\mu\mu} |^{-\frac{1}{2}} \tag{5.1}$$

when $m_{\mu i} \equiv 0$.

PROOF. We note that $\| \tau_{(i)}^{\mu\nu} \| = \| \tau_{(i)}^{\mu\mu} \delta_{\mu\nu} \|$ and hence the determinant in (4.1) consists of diagonal blocks with zeros elsewhere. When we expand, (5.1) is obtained.

COROLLARY. *Assuming the conditions of Theorem 2 and the condition that* $\| \tau_{(i)}^{\mu\nu} \|$ *has for each* i *the same principal axes and* $\| \sigma_{ij}^{(\mu)} \| = \| \sigma_{ij} \|$ *independent of* μ*, then*

$$E_{\beta_r} = \left(\prod_p C_p\right) \prod_\mu | \delta_{pq} + \sigma_{pq} \lambda_{(p)}^{\mu} |^{-\frac{1}{2}} \tag{5.2}$$

when $m_{\mu i} \equiv 0$. $\{\lambda_{(i)}^{\mu}\}$ *are the characteristic roots of the matrix* $\| \tau_{(i)}^{\mu\nu} \|$ *and the superscripts* μ *yield corresponding roots in the* k*-dimensional spaces for the* n *values of* i.

The proof is obtained by rotating each k-dimensional space to diagonalize the matrices. The same rotation will diagonalize for each i. Because $\| \sigma_{ij}^{(\mu)} \|$ is independent of μ the covariance matrix for the predictions will be unchanged.

6. Limiting distribution (number of predictions $n \to \infty$). Because the expression for E_r is a sum of $\binom{n}{r}$ terms, the numerical calculations for large values of n would be prodigious. Consequently we introduce several limiting distributions obtained by letting n increase indefinitely, subject to suitable conditions. The limiting conditions in each case should indicate the applicability in particular situations.

Concerning the question of the existence of the different limiting distributions, a sufficient condition would be the convergence of the series for $Pr\{R = r\}$,

$$Pr\{R = r\} = \sum_{s=0}^{\infty} (-1)^s \binom{r+s}{r} E_{r+s},$$

obtained by having

$$\lim_{r \to \infty} \frac{r! E_r}{m^r} \leq 1 \tag{6.1}$$

for some value of m.

When n becomes large, the enumeration of predictions is unwieldy. Therefore, they will be given in terms of time, a convenient parameter for intuitive consideration. Thus we write

$$\sigma_{ij}^{(\mu)} = \sigma_{(t_i, t_j)}^{(\mu)},$$

where this is the covariance between the μth coordinates of the predictions at times t_i and t_j. Also

$$\tau^{\mu\nu}_{(i)} = \tau^{\mu\nu}_{(t_i)},$$

an element of the success function matrix at time t_i.

THEOREM 4. TYPE I. *Assuming the conditions of Theorem 2 and letting $n \to \infty$ so that the predictions are uniformly spaced from 0 to T and the success functions approach 0 as $1/n$ with $D(t) = nC(t)$ bounded and independent of n, then*

$$(6.2) \quad E_r = \frac{1}{r!T^r}\int_0^T \cdots \int_0^T \left(\prod_1^r D(t_p)\right) |\delta_{pq}\delta_{\mu\nu} + \sigma^{(\mu)}_{(t_p,t_q)}\tau^{\mu\nu}_{(t_p)}|^{-\frac{1}{2}}\, dt_1\, dt_2 \cdots dt_r,$$

where $m_{\mu i} \equiv 0$.

PROOF. The minimum value of $|\delta_{pq}\delta_{\mu\nu} + \sigma^{(\mu)}_{(t_p,t_q)}P^{\mu\nu}_{(t_p)}|$, which is the determinant of a positive definite matrix with 1's added down the diagonal, will be greater than 1. If in addition $D(t)$ is bounded, we have

$$\lim_{r\to\infty} \frac{r!E_r}{[\sup D(t)]^r} \leq 1,$$

and this is sufficient to assure the existence of the limiting distribution.

$$E_r = \lim_{n\to\infty} \sum_{\beta_r} \left\{\left(\prod_p \frac{D(t_p)}{n}\right) \middle| \delta_{pq}\delta_{\mu\nu} + \sigma^{(\mu)}_{(t_p,t_q)}\tau^{\mu\nu}_{(t_p)} \right|^{-\frac{1}{2}}$$

$$= \lim_{n\to\infty} \frac{(1 - 1/n)\cdots(1 - r/n)}{r!}$$

$$\cdot \sum_{\substack{\text{all permutations} \\ r \text{ from } n}} \frac{(\prod_p D(t_p)\ |\ \delta_{pq}\delta_{\mu\nu} + \sigma^{(\mu)}_{(t_p,t_q)}\tau^{\mu\nu}_{(t_p)}|^{-\frac{1}{2}}}{n(n-1)\cdots,(n-r+1)}$$

$$= \frac{1}{r!T^r}\int_0^T \cdots \int_0^T \left(\prod_{p=1}^r D(t_p)\right) \left| \delta_{pq}\delta_{\mu\nu} + \sigma^{(\mu)}_{(t_p,t_q)}\tau^{\mu\nu}_{(t_p)} \right|^{-\frac{1}{2}} \prod_{p=1}^r dt_p.$$

This completes the proof.

The applicability of this distribution as an approximation for large values of n will be discussed for the Cunningham and Hynd problem in a later paper.

THEOREM 5. TYPE II. *Assuming the conditions of Theorem 2 and letting $n \to \infty$ and the scale of the success functions decrease such that $\tau^{\mu\nu}_{(t_i)} = n^{2/k}\tau'^{\mu\nu}_{(t_i)}$, then*

$$(6.3) \quad E_r = \frac{1}{r!T^r}\int_0^T \cdots \int_0^T \left(\prod_{p=1}^r C(t_p)\ |\ T'^{\mu\nu}_{(t_p)}|^{-\frac{1}{2}}\right) \prod_\mu |\ \sigma^{(\mu)}_{(t_p,t_q)}|^{-\frac{1}{2}} \prod_{p=1}^r dt_p$$

if $m_{\mu i} \equiv 0$, $|\tau'^{\mu\nu}_{(i)}|$ is bounded from 0, and

$$\int_0 \cdots \int_0^T \prod_\mu |\ \sigma^{(\mu)}_{(t_p,t_q)}|^{-\frac{1}{2}} \prod_{p=1}^r dt_p$$

exists and $\leq m^r$ *where* m *is independent of* r.

PROOF. The existence of the limiting distribution is guaranteed by these last conditions, which insure that the set $\{E_r\}$ satisfies the condition given by formula (6.1).

$$\begin{aligned} E_r &= \lim_{n\to\infty} \sum_{\beta_r} \{(\prod_p C(t_p)) \mid \delta_{pq}\delta_{\mu\nu} + \sigma^{(\mu)}_{(t_p, t_q)} n^{2/k} \tau'^{\mu\nu}_{(t_p)} \mid^{-\frac{1}{2}}\} \\ &= \lim_{n\to\infty} \sum_{\beta_r} \{(\prod_p C(t_p))(\prod_\mu \mid \sigma^{(\mu)}_{(t_p, t_q)} \mid^{-\frac{1}{2}}) \mid \sigma^{(t_p, t_q)}_{\beta_r(\mu)} \delta_{\mu\nu} + \tau'^{\mu\nu}_{(t_p)} n^{2/k} \delta_{pq} \mid^{-\frac{1}{2}}\} \\ &= \lim_{n\to\infty} \sum_{\beta_r} \left\{ \frac{(\prod_p C(t_p))(\prod_\mu \mid \sigma^{(\mu)}_{(t_p t_q)} \mid^{-\frac{1}{2}}) \prod_p \mid \tau'^{\mu\nu}_{(t_p)} \mid^{-\frac{1}{2}}(1 + O(n^{-2/k}))}{n^r} \right\} \\ &= \frac{1}{r!T^r} \int_0^T \cdots \int_0^T \prod_p (C(t_p) \mid \tau'^{\mu\nu}_{(t_p)} \mid^{-\frac{1}{2}}) \prod_\mu \mid \sigma^{(\mu)}_{(t_p, t_q)} \mid^{-\frac{1}{2}} \prod_{p=1}^{r} dt_p . \end{aligned}$$

This proves the theorem.

THEOREM 6. TYPE III. *Assuming the conditions of Theorem 2 and letting* $n \to \infty$ *with the scale of the success function increasing, and its density at any trial decreasing according to*

$$\text{(a)} \quad \tau^{\mu\nu}_{(t)} = \tau'^{\mu\nu}_{(t)} n^{-a^2},$$

$$\text{(b)} \quad C(t) = \frac{1}{n} D(t),$$

then

$$E_r = \frac{1}{r!}\left[\frac{1}{T}\int_0^T D(t)\, dt\right]^r, \tag{6.4}$$

where $m_{\mu i} \equiv 0$.

The proof of this theorem is similar to that of Theorem 4. Note that the condition for a limiting distribution is satisfied so long as $D(t)$ is bounded.

This distribution is the Poisson distribution with the usual Poisson parameter

$$m = \frac{1}{T}\int_0^T D(t)\, dt.$$

7. Limiting distributions for a fixed number of predictions n.

THEOREM 7. *When the scale of the success function increases, the distribution approaches the simple generalization of the binomial where the probability of successful predictions need not be the same for each trial, and*

$$E_r = \sum_{\beta_r} \prod_{p \in \beta_r} C_p , \tag{7.1}$$

where $m_{\mu i} \equiv 0$.

THEOREM 8. *When the correlation between the values of particular coordinates of* $\bar{X}_i$ *approaches* 0, *then the simple binomial generalization is obtained, with*

$$E_r = \sum_{\beta_r} \prod_{p \in \beta_r} \{C_p \mid \delta_{\mu\nu} + \sigma^{(\mu)}_{pp} \tau^{\mu\nu}_{(p)} \mid^{-\frac{1}{2}}\}. \tag{7.2}$$

THEOREM 9. *When the correlation between the values at different trials of particular coordinates of* $\bar{X}_i$ *approaches* 1, *then*

$$E_r = \sum_{\beta_r} \left(\prod_p C_p\right) \mid \delta_{pq}\delta_{\mu\nu} + \sqrt{\sigma_{pp}^{(\mu)}}\sqrt{\sigma_{qq}^{(\mu)}}\,\tau_{(p)}^{\mu\nu} \mid^{-\frac{1}{2}}. \tag{7.3}$$

The proofs for Theorems 7, 8, and 9 follow routine lines.

8. Time average for a fixed number of predictions *n*. If the conditions for our generalized distribution vary with time, then an expression for the probabilities obtained as a time average would be appropriate.

THEOREM 10. *Assuming the time interval between predictions is h and that the first prediction occurs at an undetermined time in the interval* (0, T'), *then the general distribution has its probabilities determined by*

$$E_r = \sum_{\beta_r} \frac{1}{T'} \int_{0-h}^{T'-h} \left(\prod_p C(t + ph)\right) \mid \delta_{pq}\delta_{\mu\nu} + \sigma_{(t+ph,\,t+qh)}^{(\mu)}\,\tau_{(t+ph)}^{\mu\nu} \mid^{-\frac{1}{2}} dt. \tag{8.1}$$

PROOF. Assuming that the time of the first prediction is uniformly distributed on the interval (0, T'), (8.1) follows from Theorem 2.

9. Acknowledgement. The author wishes to express his appreciation to Professor S. S. Wilks, under whose guidance this paper was prepared, and to thank Professor John W. Tukey for valuable discussions of the problem.

REFERENCE

[1] L. B. C. CUNNINGHAM AND W. R. B. HYND, "Random processes in air warfare," *Jour. Roy. Stat. Soc., Suppl.*, Vol. 8 (1946), p. 62–85.

ESTIMATION OF PARAMETERS IN TRUNCATED PEARSON FREQUENCY DISTRIBUTIONS

By A. C. Cohen, Jr.

The University of Georgia

1. Introduction and summary. A method based on higher moments is presented in this paper by which the type of a univariate Pearson frequency distribution (population) can be determined and its parameters estimated from truncated samples with known points of truncation and an unknown number of missing observations. Estimating equations applicable to the four-parameter distributions involve the first six moments of a doubly truncated sample or the first five moments of a singly truncated sample. When the number of parameters to be estimated is reduced, there is a corresponding reduction in the order of the sample moments required. A sample is described as singly or doubly truncated according to whether one or both "tails" are missing. Estimates obtained by the method of this paper enjoy the property of being consistent and they are relatively simple to calculate in practice. They should be satisfactory for (a) rough estimation, (b) graduation, and (c) first approximations on which to base iterations to maximum likelihood estimates.

Previous investigations of truncated univariate distributions include studies of truncated normal distributions by Pearson and Lee [1], [2], Fisher [3], Stevens [4], Cochran [5], Ipsen [6], Hald [7], and this writer [8], [9]. In addition, the truncated binomial distribution has been studied by Finney [10], and the truncated Type III distribution by this writer [11].

2. Complete distributions. The Pearson system of frequency curves has its genesis in the differential equation

$$\frac{1}{f(x)}\frac{df(x)}{dx} = \frac{a - x}{b_0 + b_1 x + b_2 x^2}, \tag{1}$$

where the origin is arbitrarily taken. Since we are concerned with truncated distributions, it is convenient to take the origin at the left terminus. In the derivations which follow we regard a, b_0, b_1, and b_2 as primary characterizing parameters of the distributions studied. The mean, standard deviation, α_3, and α_4 are expressed as functions of these quantities. To obtain a moment recursion formula for the general Pearson (complete) function, $f(x)$, we separate the variables of (1), multiply both sides of the resulting equation by x^k, and integrate over the full range of permissible values of x, i.e., $r \leq x \leq s$. Thereby we obtain

$$\int_r^s (b_0 + b_1 x + b_2 x^2) x^k \, df(x) = \int_r^s (a - x) x^k f(x) \, dx. \tag{2}$$

Let the kth moment of the complete (population) distribution function, $f(x)$, about the origin selected (i.e., about the left terminus) be designated as

$$\mu_k' = \int_r^s x^k f(x)\,dx, \tag{3}$$

and the right member of (2) becomes

$$a\mu_k' - \mu_{k+1}'.$$

Upon integrating the left member of (2) by parts, we obtain

$$[(b_0 + b_1x + b_2x^2)x^k f(x)]_r^s - kb_0\mu_{k-1}' - (k-1)b_1\mu_k' - (k+2)\mu_{k+1}'.$$

The Pearson system includes only those solutions of (1) for which $f(r) = f(s) = 0$, and moreover only those for which the left member of the above expression vanishes at both limits. As a consequence of these restrictions, we may combine the left and right members above, to obtain the following recursion formula for moments of the complete distribution about the origin:

$$h\mu_k' + b_0k\mu_{k-1}' + b_1k\mu_k' + b_2(k+2)\mu_{k+1}' = \mu_{k+1}', \tag{4}$$

where we have written

$$h = a + b_1. \tag{5}$$

If $f(x)$ is normalized so that $\mu_0' = 1$, and we let $k = 0, 1, 2$, and 3, successively, in (4), the resulting system of equations may be written as

$$\begin{aligned}
(2b_2 - 1)\mu_1' &= -h,\\
(b_1 + h)\mu_1' + (3b_2 - 1)\mu_2' &= -b_0,\\
2b_0\mu_1' + (2b_1 + h)\mu_2' + (4b_2 - 1)\mu_3' &= 0,\\
3b_0\mu_2' + (3b_1 + h)\mu_3' + (5b_2 - 1)\mu_4' &= 0.
\end{aligned} \tag{6}$$

On solving (6) for moments of the complete distribution we obtain

$$\begin{aligned}
\mu_1' &= h/(1 - 2b_2),\\
\mu_2' &= [b_0 + (b_1 + h)\mu_1']/(1 - 3b_2),\\
\mu_3' &= [2b_0\mu_1' + (2b_1 + h)\mu_2']/(1 - 4b_2),\\
\mu_4' &= [3b_0\mu_2' + (3b_1 + h)\mu_3']/(1 - 5b_2).
\end{aligned} \tag{7}$$

With h, b_0, b_1, and b_2 known, it is a simple matter to determine μ_1', μ_2', μ_3', and μ_4' from equations (7) in the order named. These equations might, of course, be rewritten with μ_1', μ_2', and μ_3' entirely eliminated from the right members. However, when this is done they become too complex in structure to be of practical value. After calculating μ_1', μ_2', μ_3', and μ_4' from (7), corresponding central moments can be determined from the well known translation formula

$$\mu_k = \sum_{i=0}^{k} \binom{k}{i} \mu_{k-i}'(\mu_1')^i, \tag{8}$$

and the standard moments from

$$\alpha_k = \mu_k/\sigma^k, \tag{9}$$

where $\sigma^2 = \mu_2$. The second central moment then becomes

$$\mu_2 = \frac{1}{1-3b_2}\left[b_0 + \frac{h}{1-2b_2}\left\{b_1 + \frac{hb_2}{1-2b_2}\right\}\right]. \tag{10}$$

Similar formulas can also be written for μ_3 and μ_4, but they are too unwieldy to be useful. For each practical application, it seems preferable to compute noncentral moments about the left terminus from (7). Central and standard moments, as required, can then be obtained from (8) and (9).

If we designate the left truncation point in standard units of the complete distribution by ξ', we have $\xi' = (0 - \mu_1')/\sigma$ and thus

$$\mu_1' = -\sigma\xi'. \tag{11}$$

Although formulas expressing the mean, standard deviation, α_3, and α_4 explicitly as functions of a, b_0, b_1, and b_2 are unduly complex for the four-parameter distributions, as shown below they become relatively simple for Type III and Normal distributions.

Type III distribution. In this case $b_2 = 0$, and we have

$$\begin{aligned} \mu_1' &= h, & h &= -\sigma\xi', \\ \sigma &= \sqrt{b_0 + b_1 h}, & b_1 &= \sigma\alpha_3/2, \\ \alpha_3 &= 2b_1/\sqrt{b_0 + b_1 h}, & b_0 &= \sigma^2[1 + \xi'\alpha_3/2]. \end{aligned} \tag{12}$$

Normal distribution. In this case $b_1 = b_2 = 0$, and

$$\begin{aligned} \mu_1' &= h, & h &= -\sigma\xi', \\ \sigma &= \sqrt{b_0}, & b_0 &= \sigma^2. \end{aligned} \tag{13}$$

3. Recursion formula for moments of incomplete distributions. If the limits of integration in equation (2) are reduced to include only the truncated range $0 \le x \le d$, where $r \le 0$ and $d \le s$, we have

$$\int_0^d (b_0 + b_1 x + b_2 x^2)x^k\, df(x) = \int_0^d (a - x)x^k f(x)\, dx. \tag{14}$$

Define the kth moment of the truncated distribution about the left terminus as

$$m_k = \int_0^d x^k f(x)\, dx, \tag{15}$$

with $m_0 = 1$, and the right member of (14) becomes

$$am_k - m_{k+1}.$$

On integrating the left member of (14) by parts we obtain

$$[(b_0 + b_1 x + b_2 x^2)x^k f(x)]_0^d - kb_0 m_{k-1} - (k+1)b_1 m_k.$$

Since we are not integrating over the full range of x, the first term of the above expression does not vanish as it did with the complete distribution. However, if we define

$$F_1 = f(0)b_0, \quad F = f(d)[b_0 + db_1 + d^2b_2], \tag{16}$$

and then combine left and right members above, we obtain the following recursion formula for moments of the truncated distribution:

$$hm_k + b_0km_{k-1} + b_1km_k + b_2(k+2)m_{k+1} - d^kF = m_{k+1} \qquad (k \geq 1). \tag{17}$$

If we let $k = 0$ in (14) prior to integrating, and then proceed as outlined above, we obtain

$$h + 2m_1b_2 + F_1 - F = m_1, \tag{18}$$

which may be regarded as a companion equation to (17) for the case $k = 0$.

4. Estimating h, b_0, b_1, and b_2 from doubly truncated samples. To obtain estimates by equating observed sample moments to population moments, we substitute the observed sample moments ν_k for the m_k in (17), simultaneously replacing the population parameters h, b_j, $\cdots$, and F by their estimates h^*, b_j^*, $\cdots$, F^*. Setting $k = 1, 2, \cdots, 5$, successively, we find the estimating equations

$$\begin{aligned}
\nu_1h^* + b_0^* + \nu_1b_1^* + 3\nu_2b_2^* - dF^* &= \nu_2, \\
\nu_2h^* + 2\nu_1b_0^* + 2\nu_2b_1^* + 4\nu_3b_2^* - d^2F^* &= \nu_3, \\
\nu_3h^* + 3\nu_2b_0^* + 3\nu_3b_1^* + 5\nu_4b_2^* - d^3F^* &= \nu_4, \\
\nu_4h^* + 4\nu_3b_0^* + 4\nu_4b_1^* + 6\nu_5b_2^* - d^4F^* &= \nu_5, \\
\nu_5h^* + 5\nu_4b_0^* + 5\nu_5b_1^* + 7\nu_6b_2^* - d^5F^* &= \nu_6.
\end{aligned} \tag{19}$$

These constitute a linear system of five equations in the five estimates, h^*, b_0^*, b_1^*, b_2^*, and F^*, which can be solved by any of the standard methods applicable to such systems. For practical applications, the writer suggests using either the method of "single division" or "multiplication and subtraction" as described by Dwyer [12]. With estimates h^*, b_0^*, b_1^*, and b_2^* thus calculated, we substitute these values in (7) to estimate moments of the complete distribution, and subsequently compute the required estimates of population (complete distribution) parameters with the aid of (8) and (9). F_1^* can be computed from (18) upon replacing parameters by their estimates and m_1 by ν_1. It will be noted that these estimates are consistent since if they should be calculated from the entire population they would obviously equal the required parameters.

Although neither F_1^* nor F^* is required in estimating moments of the complete distribution, a comparison of their values found on solving (18) and (19) with corresponding values computed from the finally fitted curve with the

aid of (16) affords a check on agreement between the fitted curve and observed sample data.

It should be noted here that estimates are distinguished from parameters throughout this paper by starring (*) the estimates.

5. Determining type of distribution. With estimates of μ_1', σ, α_3, and α_4 computed as indicated in Section 4, the type of the distribution involved can be established from the original Pearson criteria, an excellent exposition of which has been given by Elderton [13], or from the Carver-Craig criteria [14]. In the present instance, however, since estimates of b_0, b_1, and b_2 must of necessity be computed before estimates of the population parameters can be obtained, it seems more appropriate to determine the type directly from the quadratic equation

$$b_0 + b_1x + b_2x^2 = 0. \tag{20}$$

The general solution of the differential equation (1) can be written as

$$f(x) = C(x - r_1)^{m_1}(r_2 - x)^{m_2}, \tag{21}$$

where r_1 and r_2 are roots of (20) (cf., for example, [14]). The nature of these roots determines the type of the distribution. If we let D designate the discriminant, $D = b_1^2 - 4b_0b_2$, the principal Pearson curves[1] may be classified as follows:

Type I	$r_1 - \mu_1' < 0 < r_2 - \mu_1'$,	$D > 0$;
II	$(r_1 - \mu_1') = -(r_2 - \mu_1')$, $b_1 = 2b_2\mu_1'$,	$D > 0$;
III	$b_2 = 0$	
IV	r_1 and r_2 imaginary, $b_1 \neq 2b_2\mu_1'$,	$D < 0$;
V	$(r_1 - \mu_1')$, $(r_2 - \mu_1')$ of the same sign,	$D > 0$;
VI	r_1 and r_2 imaginary, $b_1 = 2b_2\mu_1'$,	$D < 0$;
Normal	$b_1 = b_2 = 0$.	

It can be shown that a necessary condition for the odd central moments to equal zero (i.e., for $f(x)$ to be symmetrical about is mean) is that

$$b_1 = 2b_2\mu_1'.$$

6. Singly truncated samples. If only the left tail is omitted, then F vanishes, and we can drop from (19) the last equation, which would otherwise be required, after placing $F^* = 0$ in the remaining equations. If only the right tail is missing, then $F_1 = 0$, and by changing the variable from x to $d - x$ we can translate the origin to the truncation point on the right, set $F_1^* = 0$, and again drop the last equation otherwise required in (19). As an alternate and frequently preferable

[1] The numbering of the types followed here is that of Craig [14].

procedure when some origin other than the truncation point of a singly truncated sample has been employed, we might substitute (18) for the last equation of (19) after replacing parameters by their estimates. In both instances, the order of the highest order sample moment required is reduced by one from the requirements for doubly truncated samples.

In practical applications, finding either F_1^* or F^* equal or almost equal to zero from a sample that is represented as being doubly truncated, suggests that perhaps the sample was in fact only singly truncated. In this case, either the sample terminus is the terminus of the complete distribution or the absence of lower sample values is due to the small probability associated with their occurrence. Finding both F_1^* and F^* equal or nearly equal to zero suggests that the sample was not truncated after all, and that the necessary estimates should be computed from estimating equations applicable to complete samples.

When the sample terminus is employed as an estimate of the corresponding population terminus, an additional equation may be dropped from (19) since in this case we are estimating one less parameter from the moment equations. To illustrate, consider a Type III distribution for which the left sample terminus (origin) is considered as an appropriate estimate of the population lower limit. We then have

$$h = 2\sigma/\alpha_3,$$

and from (12)

$$h = (b_0 + b_1 h)/b_1.$$

Consequently it follows that $b_0 = 0$, and the system of estimating equations to be solved consists of the first two equations of (19) plus (18) with $b_0^* = b_2^* = 0$. The parameters appearing in (18) are of course replaced by their estimates.

7. Type III and normal distributions. When it is desired to estimate parameters of a Type III distribution (for which $b_2 = 0$) from a doubly truncated sample, we need calculate only the first five sample moments and solve the first four equations of (19) after placing $b_2^* = 0$. With singly truncated samples from which the left tail is missing, we require only the first four sample moments and need solve only the first three equations of (19) after setting $F^* = 0$.

To estimate parameters of a normal distribution (for which $b_1 = b_2 = 0$) from doubly truncated samples, we calculate the first four sample moments and solve the first three equations of (19) after setting $b_1^* = b_2^* = 0$. With singly truncated samples from which the left tail is missing, we require only the first three sample moments and need solve only the first two equations of (19) after setting $F^* = 0$.

8. A numerical example. To illustrate the application of results obtained in this paper to practical problems, we consider an example given by Miss Shook [15] on the weights of 1000 female students (cf. Table 1). Miss Shook considered her data as a complete (untruncated) sample from a Pearson Type III popula-

TABLE 1

Weights of 1000 female students

Weight in pounds	Observed frequency	Graduated frequencies based on Type III distribution: Complete sample	Truncated on right	Limit at sample terminus	Doubly truncated
70– 79.9	2	0	0.2	0.0	0.2
80– 89.9	16	4	12.8	8.4	12.7
90– 99.9	82	102	94.0	97.5	94.0
100–109.9	231	238	213.7	223.6	214.0
110–119.9	248	250	254.1	247.9	253.9
120–129.9	196	184	200.9	191.7	200.8
130–139.9	122	111	120.7	118.6	120.6
140–149.9	63	59	59.5	63.2	59.4
150–159.9	23	29	25.2	30.2	25.2
160–169.9	5	13	9.5	13.3	9.5
170–179.9	7	6	3.3	5.5	3.3
180–189.9	1	3	1.0	2.2	1.0
190–199.9	2	1	0.3	0.8	0.3
200–209.9	1	0	0.1	0.0	0.1
210–219.9	1	0	0.0	0.0	0.0
Total.............	1000	1000	995.3	1002.9	995.0
Total frequency in truncated range..	981	977	980.9	981.1	980.6
M^* (lbs.)		118.74	118.55	119.14	118.56
σ^* (lbs.)		16.9175	16.027	16.958	16.024
α_3^*		0.976424	0.655	0.865	0.657
Lower limit (lbs.)..........		84.09	69.61	79.95	69.77
F_1^* (from sample moments).............			0	0	−0.002
F_1^* (from fitted curve)..................			0.006	0	0.005
F^* (from sample moments).............			0.767	1.358	0.769
F^* (from fitted curve)..................			0.732	1.183	0.733

Truncated sample obtained by truncating the complete sample on the left at 79.95 lbs. and on the right at 159.95 lbs.

tion, and employed the method of moments to estimate population parameters. Using these estimates, she then graduated the observed sample data.

For our purposes, we truncate Miss Shook's sample on the left at 79.95 lbs. and on the right at 159.95 lbs., thus eliminating the first and the last six cells of the grouped data. The retained (truncated) sample then consists of 981 observations, all of which are within the range 79.95 to 159.95 lbs. We disregard all prior knowledge about the type of the population, and accordingly compute the first six sample moments about the lower terminus. In order to compensate for moment errors due to grouping, we apply Sheppard's[2] corrections for non-central moments. Both sets of moments are given below.

Uncorrected moments	*Corrected moments*
$n_1 = (7.56676860)\ 5$	$\nu_1 = (7.56676860)\ 5$
$n_2 = (66.4026504)\ 5^2$	$\nu_2 = (66.0693171)\ 5^2$
$n_3 = (649.817533)\ 5^3$	$\nu_3 = (642.250764)\ 5^3$
$n_4 = (6913.71764)\ 5^4$	$\nu_4 = (6781.37901)\ 5^4$
$n_5 = (78479.9827)\ 5^5$	$\nu_5 = (76331.5834)\ 5^5$
$n_6 = (937015.638)\ 5^6$	$\nu_6 = (902910.393)\ 5^6$

We substitute these values in (19) and solve the system by Dwyer's method of multiplication and subtraction to obtain $h^* = 44.973178$, $b_0^* = -53.5929$, $b_1^* = 12.339508$, $b_2^* = -0.084107$, and $F^* = 0.578321$. From (18) we then obtain $F_1^* = -0.196817$. The small negative value thus computed suggests that perhaps F_1 actually has the value zero and that there was no truncation on the left.

Considering the sample as being truncated on the right only, and rather than translate our origin to the right sample terminus, we substitute (18) with $F_1^* = 0$ for the last equation of (19) to obtain a new system of five equations in the same five unknown estimates as before, but involving only the first five sample moments. On solving the new set of equations, we obtain $h^* = 38.530928$, $b_0^* = 54.83444$, $b_1^* = 5.179891$, $b_2^* = 0.000986$, and $F^* = 0.771707$.

The small values obtained for b_2^* in both the above cases lead us to conjecture that b_2 actually has the value zero, and that our sample came from a Type III population.

With the sample considered as coming from a Type III population and as being truncated on the right only, we solve the system consisting of the first three equations of (19) plus (18) with $b_2^* = F_1^* = 0$, and obtain $h^* = 38.600670$, $b_0^* = 54.1194$, $b_1^* = 5.247727$, and $F^* = 0.766827$. On substituting these values in (12) we have $\mu_1'^* = 38.60$ lbs., $\sigma^* = 16.027$ lbs., and $\alpha_3^* = 0.655$. The mean referred to zero as an origin is estimated as $M^* = \mu_1'^* + 79.95$ lbs. $= 118.55$ lbs. The corresponding estimate of the lower limit is 69.61 lbs. A graduation of the sample data using these estimates and carried out with the aid of Salvosa's

[2] See for example reference [16], formula 27.9.3, page 361.

tables [17] is given in Table 1, along with Miss Shook's original graduation which was based on estimates from the complete sample.

To provide additional comparisons, we compute further estimates with the sample assumed to be doubly truncated from a Type III population. Accordingly, we solve the first four equations of (19) with $b_2^* = 0$ to obtain $h^* = 38.605540$, $b_0^* = 53.5835$, $b_1^* = 5.262710$, and $F^* = 0.769439$. From (18) we find $F_1^* = -0.002258$. Similarly, we calculate an additional set of estimates under the assumption that the sample was truncated on the right only but with the left sample terminus being the lower limit of the complete distribution. In this case, the system of three equations consisting of the first two equations of (19) plus (18) with $b_0^* = b_2^* = 0$ yields the solutions $h^* = 39.191957$, $b_1^* = 7.337336$, and $F^* = 1.358114$. Estimates of the basic population parameters for each of the above cases, along with graduations over the complete sample range, are also included in Table 1.

The agreement between observed and graduated frequencies is found to be much better for estimates based on the truncated sample than for estimates based on the complete sample. The improved results obtained with the truncated sample suggest that perhaps some of the extreme observations in Miss Shook's original data came from a different population than that which accounted for the main body of her data. It makes little difference whether the truncated sample is considered as being singly or doubly truncated or whether the left sample terminus is used as an estimate of the population lower limit or not. It will be also noted that the values of F_1^* and F^* as computed from the finally fitted curves with the aid of (16) are in substantially close agreement with the corresponding values found on solving the moment equations. In each case the graduations are very nearly equal throughout the entire sample range, and any one of the truncated sample graduations would be considered as a satisfactory fit to the observed data over the truncated range. Certainly any one of the three sets of estimates would, for this example, be adequate as first approximations on which to base iterations to maximum likelihood estimates in a manner similar to that previously employed by Koshal [18] in improving moment estimates from complete samples. The writer hopes to give further consideration in a subsequent paper to the problems of such iterations when samples are truncated.

REFERENCES

[1] KARL PEARSON AND ALICE LEE, "On the generalized probable error in multiple normal correlation," *Biometrika*, Vol. 6 (1908), pp. 59–68.

[2] ALICE LEE, "Table of Gaussian 'tail' functions when the 'tail' is larger than the body," *Biometrika*, Vol. 10 (1915), pp. 208–215.

[3] R. A. FISHER, "Properties and applications of H_h functions," *Mathematical Tables*, Vol. 1, pp. xxvi–xxxv, British Association for the Advancement of Science, 1931.

[4] W. L. STEVENS, "The truncated normal distribution," Appendix to "The calculation of the time-mortality curve" by C. I. Bliss, *Annals of Applied Biology*, Vol. 24 (1937), pp. 815–852.

[5] W. G. Cochran, "Use of IBM equipment in an investigation of the truncated normal problem," *Proc. Research Forum*, International Business Machines Corp., 1946, pp. 40–43.

[6] Johannes Ipsen, Jr., "A practical method of estimating the mean and standard deviation of truncated normal distributions," *Human Biology*, Vol. 21 (1949), pp. 1–16.

[7] A. Hald, "Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point," *Skandinavisk Aktuarietidskrift*, Haft 3-4 (1949), pp. 119–134.

[8] A. C. Cohen, Jr., "On estimating the mean and standard deviation of truncated normal distributions," *Jour. Am. Stat. Assn.*, Vol. 44 (1949), pp. 518–525.

[9] A. C. Cohen, Jr., "Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 557–569.

[10] D. J. Finney, "The truncated binomial distribution," *Annals of Eugenics*, Vol. 14 (1949), pp. 319–328.

[11] A. C. Cohen, Jr., "Estimating parameters of Pearson Type III populations from truncated samples," *Jour. Am. Stat. Assn.*, Vol. 45 (1950), pp. 411–423.

[12] P. S. Dwyer, "The solution of simultaneous equations," *Psychometrika*, Vol. 6 (1941), pp. 101–129.

[13] W. P. Elderton, *Frequency Curves and Correlation*, 3rd ed., Cambridge University Press, 1938.

[14] C. C. Craig, "A new exposition and chart for the Pearson system of frequency curves," *Annals of Math. Stat.*, Vol. 7 (1936), pp. 16–28.

[15] B. L. Shook, "Synopsis of elementary mathematical statistics," *Annals of Math. Stat.*, Vol. 1 (1930), pp. 14–41.

[16] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[17] Luis R. Salvosa, "Tables of Pearson's Type III function," *Annals of Math. Stat.*, Vol. 1 (1930), appended.

[18] R. S. Koshal, "Application of the method of maximum likelihood to the improvement of curves fitted by the method of moments," *Jour. Roy. Stat. Soc.*, Vol. 96 (1933), pp. 303–13.

ON THE DISTRIBUTION OF THE CHARACTERISTIC ROOTS OF NORMAL SECOND-MOMENT MATRICES[1]

By A. M. Mood

The Rand Corporation

1. Summary. Distributions of characteristic roots have been obtained by Girshick [1], Fisher [2], Hsu [3], and Roy [4]. The present paper outlines an alternative derivation of these distributions which is somewhat more elementary than those that have been published and which may have some pedagogical utility. The primary object of the paper, however, is to obtain the normalizing constants for these distributions; though the correct values of the constants have been published in the references cited above, no convincing derivation seems to have been recorded.

2. The problem. Let

$$a_{ij} = \sum_{\alpha=1}^{m} (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j) \qquad (i, j = 1, 2, \cdots, k) \tag{1}$$

be sums of squares and products for samples of size $m(>k)$ from a k-variate normal distribution with covariance matrix $\| \sigma_{ij} \| (= \| \sigma^{ij} \|^{-1})$ having a k-fold characteristic root λ. The a_{ij} are distributed by the Wishart density function

$$f(a_{ij}\,; m - 1, \sigma^{ij}) = \frac{\left| \tfrac{1}{2} \sigma^{ij} \right|^{\frac{1}{2}(m-1)} \left| a_{ij} \right|^{\frac{1}{2}(m-k-2)} e^{-\frac{1}{2}\Sigma \sigma^{ij} a_{ij}}}{\pi^{\frac{1}{4}k(k-1)} \prod_{i=1}^{k} \Gamma[\tfrac{1}{2}(m - i)]} \tag{2}$$

with $m - 1$ degrees of freedom. Let b_{ij} be similarly distributed with $n - 1$ degrees of freedom and independently of the a_{ij}.

We are concerned with the distribution of the roots $w_1, \cdots, w_k$ of

$$\left| a_{ij} - w\sigma_{ij} \right| = 0, \tag{3}$$

which roots form a natural multivariate analogue of chi-square. Similarly the roots of

$$\left| a_{ij} - vb_{ij} \right| = 0 \tag{4}$$

provide an analogue for the variance ratio, and the roots of

$$\left| a_{ij} - u(a_{ij} + b_{ij}) \right| = 0 \tag{5}$$

an analogue for the intraclass correlation. More important, the roots of (5)

---

[1] This work was done during the academic year 1939–40 when the author was a graduate student at Princeton University; it was completed just as the Hsu and Fisher papers appeared, and was therefore never submitted for publication. Recently the author learned from Hotelling that a derivation of the normalizing constants would be of interest.

are directly related to Hotelling's canonical correlations [5] for two sets of variates. For all these problems it is necessary only to obtain the distribution for the roots of (5) since the roots of (4) are

$$v_i = u_i/(1 - u_i) \tag{6}$$

and the distribution of the w_i may be obtained by letting $n \to \infty$ in the distribution of the u_i.

3. Density function for the u_i. It is no essential simplification to suppose, as we shall do, that

$$\begin{aligned} \sigma_{ij} = \delta_{ij} &= 1 \quad \text{for} \quad i = j, \\ &= 0 \quad \text{for} \quad i \neq j. \end{aligned} \tag{7}$$

The joint density for a_{ij} and b_{ij} is

$$f(a_{ij}, m - 1, \sigma^{ij})f(b_{ij}, n - 1, \sigma^{ij}), \tag{8}$$

where f is defined by (2). If $u_i(u_1 < u_2 < \cdots < u_k)$ are the roots of (5), there exists [6] a nonsingular linear transformation $\| q^{ij} \|$ such that

$$\| q^{ij} \|' \| a_{ij} + b_{ij} \| \| q^{ij} \| = \| \delta_{ij} \|, \tag{9}$$

$$\| q^{ij} \|' \| a^{ij} \| \| q^{ij} \| = \| u_i\delta_{ij} \|, \tag{10}$$

$$\| q^{ij} \|' \| b_{ij} \| \| q^{ij} \| = \|(1 - u_i)\delta_{ij} \|, \tag{11}$$

where the prime denotes the transpose.

We shall transform the $k^2 + k$ variates a_{ij} and b_{ij} of (8) to the $k^2 + k$ variates q_{ij} and u_i where

$$\| q_{ij} \| = \| q^{ij} \|^{-1}.$$

The transformed density is

$$D_1(u_i, q_{ij}) = K_1 F(q_{ij}) \left[\prod_1^k u_i\right]^{\frac{1}{2}(m-k-2)} \left[\prod_1^k (1 - u_i)\right]^{\frac{1}{2}(n-k-2)} J, \tag{12}$$

where J is the Jacobian $\partial(a_{ij}, b_{ij})/\partial(u_i, q_{ij})$ and K_1 is the normalizing constant.

We next show that J factors into a function of q_{ij} only and a function of u_i only. Let the earlier variables be ordered

$$a_{11}, a_{12}, \cdots, a_{1k}, a_{22}, a_{23}, \cdots, a_{2k}, a_{33}, \cdots, a_{3k}, \cdots, a_{kk},$$
$$b_{11}, \cdots, b_{1k}, b_{22}, \cdots, b_{2k}, \cdots, b_{kk},$$

and the new variables will be ordered

$$u_1, u_2, \cdots, u_k, q_{11}, q_{21}, \cdots, q_{k1}, q_{12}, q_{22}, \cdots, q_{k2}, \cdots, q_{1k}, \cdots, q_{kk}.$$

On differentiating the relations

$$a_{ij} = \sum_r q_{ir}q_{jr}u_r , \tag{13}$$

$$b_{ij} = \sum_r q_{ir}q_{jr}(1 - u_r), \tag{14}$$

the Jacobian can be written down directly. Supposing this to have been done (with the a_{ij} and b_{ij} corresponding to columns and the u_i and q_{ij} to rows), the result can be simplified by adding the first column of the left half to the first column of the right half, the second column of the left half to the second column of the right half, etc. The first row of the resulting determinant then has elements

$$q_{11}^2 , q_{11}q_{21} , q_{11}q_{31} , \cdots , q_{11}q_{k1} , q_{21}^2 , q_{21}q_{31} , \cdots , q_{21}q_{k1} , \cdots , q_{k1}^2$$

in the left half, and zeros in the right half. The $(k + 1)$th row, for example, has elements

$$2q_{11}u_1 , q_{21}u_1 , \cdots , q_{k1}u_1 , 0, 0, \cdots , 0$$

in the left half, and the same set with the u_1's omitted in the right half.

Now we show that J vanishes if $u_1 = u_2$. It will be easy to follow the argument if one writes down the complete Jacobian for $k = 3$. Assuming $u_1 = u_2$, the following steps produce a row of zeros in J:

1) Multiply the columns of the right half by u_1 and subtract from the corresponding columns of the left half. This makes the elements of the left half of rows $k + 1$ through $3k$ all zero.
2) Make all elements of the $(k + 1)$th row zero except the element $2q_{11}$ (in the b_{11} column) by subtracting proper multiples of the b_{11} column from the columns having nonzero elements in that row.
3) Make all elements of the $(k + 2)$th row zero except that in the b_{12} column.
4) Make all elements of the $(k + 3)$th row zero except that in the b_{13} column.

....

....

....

$k + 1$) Make all elements of the $(2k)$th row zero except that in the b_{1k} column.

$k + 2$) Make all elements of the $(2k + 1)$th row zero by subtracting proper multiples of the k rows above it from that row.

It follows therefore that J has the factor $(u_2 - u_1)$.

Similarly J must have all factors of the form $u_i - u_j$; hence J has the factor

$$\prod_{i>j} (u_i - u_j), \tag{15}$$

and since J is of total degree $k(k - 1)/2$ in the u's the other factor of J must involve only the q's.

Thus it follows that (12) factors into a function of the q_{ij} only and a function of the u_i only, say

$$D_2(u_i) = K_2 \left[\prod_1^k u_i\right]^{\frac{1}{2}(m-k-2)} \left[\prod_1^k (1 - u_i)\right]^{\frac{1}{2}(n-k-2)} \prod_{i>j} (u_i - u_j). \tag{16}$$

4. Normalizing constant. Let us define

$$L(\alpha, \beta) = \int_0^1 \int_0^{u_k} \cdots \int_0^{u_2} [\prod u_i]^\alpha [\prod (1 - u_i)]^\beta [\prod_{i>j} (u_i - u_j)] \prod du_i; \tag{17}$$

then the normalizing constant of (16) is

$$K_2 = 1/L[\tfrac{1}{2}(m - k - 2), \tfrac{1}{2}(n - k - 2)]. \tag{18}$$

Our procedure will be to first express $L(\alpha, \beta)$ as a multiple of $L(0, 0)$ and then to evaluate the latter factor directly.

In view of (9), (10), and (11),

$$\Pi u_i = |a_{ij}|/|a_{ij} + b_{ij}|, \tag{19}$$

$$\Pi(1 - u_i) = |b_{ij}|/|a_{ij} + b_{ij}|; \tag{20}$$

hence

$$E\left(\frac{|a_{ij}|^r |b_{ij}|^s}{|a_{ij} + b_{ij}|^{r+s}}\right) = \frac{L[\tfrac{1}{2}(m - k - 2) + r, \tfrac{1}{2}(n - k - 2) + s]}{L[\tfrac{1}{2}(m - k - 2), \tfrac{1}{2}(n - k - 2)]}. \tag{21}$$

But this quantity is determinable from (8) by a method due to Wilks [7]. Since the elements of $\| a_{ij} + b_{ij} \|$ are distributed by

$$f(a_{ij} + b_{ij}, m + n - 2, \sigma^{ij}), \tag{22}$$

we find first that

$$E(|a_{ij} + b_{ij}|^c) = |\tfrac{1}{2}\sigma^{ij}|^{-c} \prod_{i=1}^k \frac{\Gamma[\tfrac{1}{2}(m + n - 1 - i) + c]}{\Gamma[\tfrac{1}{2}(m + n - 1 - i)]}, \tag{23}$$

as does Wilks in [7]. Thus

$$\begin{aligned} &\int \cdots \int |a_{ij} + b_{ij}|^c f(a_{ij}, m - 1, \sigma^{ij}) f(b_{ij}, n - 1, \sigma^{ij}) \Pi da_{ij} \Pi db_{ij} \\ &\qquad = |\tfrac{1}{2}\sigma^{ij}|^{-c} \prod_1^k \frac{\Gamma[\tfrac{1}{2}(m + n - 1 - i) + c]}{\Gamma[\tfrac{1}{2}(m + n - 1 - i)]} \end{aligned} \tag{24}$$

or in another form

$$\begin{aligned} &\int \cdots \int |a_{ij} + b_{ij}|^c |a_{ij}|^{\frac{1}{2}(m-k-2)} |b_{ij}|^{\frac{1}{2}(n-k-2)} e^{[-\frac{1}{2}\Sigma \sigma^{ij}(a_{ij}+b_{ij})]} \Pi da_{ij} \Pi db_{ij} \\ &\qquad = \frac{\pi^{\frac{1}{4}k(k-1)}}{|\tfrac{1}{2}\sigma^{ij}|^{\frac{1}{2}(m+n-2)+c}} \prod_{i=1}^k \frac{\Gamma[\tfrac{1}{2}(m - i)]\Gamma[\tfrac{1}{2}(n - i)]\Gamma[\tfrac{1}{2}(m + n - 1 - i) + c]}{\Gamma[\tfrac{1}{2}(m + n - 1 - i)]}. \end{aligned} \tag{25}$$

In this expression we replace m by $m + 2r$ and n by $n + 2s$ and multiply the whole by

$$\frac{\left| \tfrac{1}{2}\sigma^{ij} \right|^{\frac{1}{2}(m+n-2)}}{\pi^{\frac{1}{2}k(k-1)} \prod_{i=1}^{k} \Gamma[\tfrac{1}{2}(m-i)]\Gamma[\tfrac{1}{2}(n-i)]}$$

to get

$$\begin{aligned} E(| a_{ij} + b_{ij} |^c \, | a_{ij} |^r \, | b_{ij} |^s) &= \left| \tfrac{1}{2}\sigma^{ij} \right|^{-c-r-s} \\ \cdot \prod_{i=1}^{k} &\frac{\Gamma[\frac{1}{2}(m-i)+r]\Gamma[\frac{1}{2}(n-i)+s]\Gamma[\frac{1}{2}(m+n-1-i)+c+r+s]}{\Gamma[\frac{1}{2}(m+n-1-i)+r+s]\Gamma[\frac{1}{2}(m-i)]\Gamma[\frac{1}{2}(n-i)]}. \end{aligned} \tag{26}$$

In this we put $c = -(r + s)$ to get an expression for the right side of (21). In the resulting relation we put $m = n = k + 2$ to get

$$L(r, s) = L(0,0)\prod_{i=1}^{k} \frac{\Gamma[\frac{1}{2}(k+2-i)+r]\Gamma[\frac{1}{2}(k+2-i)+s]\Gamma[\frac{1}{2}(2k+3-i)]}{\Gamma[\frac{1}{2}(2k+3-i)+r+s]\Gamma[\frac{1}{2}(k+2-i)]\Gamma[\frac{1}{2}(k+2-i)]}. \tag{27}$$

Now we are left only with the problem of evaluating

$$L(0, 0) = \int \underset{R}{\cdots} \int \prod_{i>j} (u_i - u_j) \prod du_i, \tag{28}$$

where R is the region $0 \leq u_1 \leq u_2 \leq \cdots \leq u_k \leq 1$. We first observe that the integrand may be put in the determinantal form

$$\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ u_1 & u_2 & u_3 & \cdots & u_k \\ u_1^2 & u_2^2 & u_3^2 & \cdots & u_k^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ u_1^{k-1} & u_2^{k-1} & u_3^{k-1} & \cdots & u_k^{k-1} \end{vmatrix}. \tag{29}$$

Thus the integrand may be written

$$\prod_{i>j} (u_i - u_j) = \sum_{\text{per}} (-1)^{t(\text{per})} \prod_{i=1}^{k} u_i^{\alpha_i - 1}, \tag{30}$$

where $\alpha_1, \alpha_2, \cdots, \alpha_k$ is a permutation of $1, 2, \cdots k$; where the sum is over all permutations of these integers; and where t(per) is the number of transpositions in the permutation.

On integrating (30) over R it is found that

$$L(0, 0) = \sum_{\text{per}} \frac{(-1)^{t(\text{per})}}{\alpha_1(\alpha_1 + \alpha_2)(\alpha_1 + \alpha_2 + \alpha_3) \cdots (\alpha_1 + \alpha_2 + \cdots + \alpha_k)}. \tag{31}$$

It is shown in the Appendix that this sum has the value

$$(32)\quad L(0, 0) = \frac{1}{k!}\prod_{i>j}\frac{i-j}{i+j}$$

$$(33)\quad = \frac{1}{k!}\frac{(k-1)!(k-2)!\cdots 2!1!}{[3\cdot 4\cdot 5\cdots(k+1)][5\cdot 6\cdots(k+2)][7\cdot 8\cdots(k+3)]\cdots[2k-1]}$$

$$(34)\quad = \prod_{i=1}^{k}\frac{\Gamma(k-1+i)\Gamma(2k+1-2i)}{\Gamma(2k+1-i)}.$$

This may also be put in the form

$$(35)\quad L(0, 0) = \frac{1}{\pi^{k/2}}\prod_{i=1}^{k}\frac{\Gamma[\frac{1}{2}(k+2-i)]\Gamma[\frac{1}{2}(k+2-i)]\Gamma[\frac{1}{2}(k+1-i)]}{\Gamma[\frac{1}{2}(2k+3-i)]}.$$

The identity of (34) and (35) is easily shown by induction on k employing the relation

$$(36)\quad \Gamma(h+1)\Gamma(h+\tfrac{1}{2}) = \sqrt{\pi}\,\Gamma(2h+1)/2^{2h}.$$

The form (35) simplifies the final expression for K_2 which is found by putting (27) and (35) in (18) to get

$$(37)\quad K_2 = \pi^{k/2}\prod_{i=1}^{k}\frac{\Gamma[\frac{1}{2}(m+n-1-i)]}{\Gamma[\frac{1}{2}(m-i)]\Gamma[\frac{1}{2}(n-i)]\Gamma[\frac{1}{2}(k+1-i)]}.$$

Putting (37) in (16) we have the density function for the roots of (5), and the densities for the roots of (3) and (4) can then be obtained as stated at the end of Section 2.

Appendix

We wish to demonstrate that if $g_i (i = 1, 2, \cdots, k)$ are k distinct positive quantities indexed in order of magnitude, then

$$(a)\quad \sum_{\text{per}}\frac{(-1)^{t(\text{per})}}{y_1(y_1+y_2)\cdots(y_1+y_2+\cdots+y_k)} \equiv \left(\prod_{i>j}\frac{g_i-g_j}{g_i+g_j}\right)\Big/\prod_{1}^{k} g_i,$$

where $y_1, y_2, \cdots, y_k$ is a permutation of $g_1, g_2, \cdots, g_k$, the sum is taken over all permutations of the g_i, and $t(\text{per})$ is the number of transpositions in the permutation $y_1, \cdots, y_k$. This identity was first formulated and proved for $g_i = i$ with considerable aid from J. B. Rosser. Here we give a different and easier argument which handles the more general situation.

First we obtain another identity as a lemma, namely,

$$(b)\quad \sum_{i=1}^{k} g_i\prod_{j\neq i}\frac{g_i+g_j}{g_i-g_j} \equiv \sum_{i=1}^{k} g_i.$$

The following argument for (b) was formulated by John Nash. The left side of (b) is a rational function of g_1, say $P(g_1)/Q(g_1)$, which we may suppose to be

reduced to its lowest terms. We first argue that the rational function is really a polynomial becuase it does not become infinite for any finite value of g_1. Certainly the only possible roots of $Q(g_1)$ are $g_2, \cdots, g_k$. Suppose $g_1 = g_2 + \epsilon$, then the first two terms (the others do not have $g_1 - g_2$ in their denominators) on the left of (b) may be written

$$\frac{2g_2 + \epsilon}{\epsilon}\left[(g_2 + \epsilon)\prod_3^k \frac{g_2 + \epsilon + g_j}{g_2 + \epsilon - g_j} - g_2 \prod_3^k \frac{g_2 + g_j}{g_2 - g_j}\right],$$

which is clearly bounded as $\epsilon \to 0$. Similarly no other g_i is a root of $Q(g_1)$; hence the left side of (b) is a polynomial $P(g_1)$ in g_1. Now let g_1 become large; the first term on the left of (b) becomes essentially g_1 while the others become constants; hence

$$P(g_1) = g_1 + C_1(g_2, \cdots, g_k).$$

Similarly as a function of g_2 the left of (b) is of the form

$$g_2 + C_2(g_1, g_3, \cdots, g_k),$$

and so forth. Furthermore, the left of (b) is homogeneous of degree one in the g's; it must therefore be $\sum_1^k g_i$.

Having (b) we can prove (a) by induction. It is true for $k = 2$, and we shall show it to be true for $k + 1$ given it to be true for k. Applying (a) to the left side of the following relation, we have

$$\sum_{\text{per}} \frac{(-1)^{t(\text{per})}}{y_1(y_1 + y_2)(y_1 + y_2 + y_3) \cdots (y_1 + y_2 + \cdots + y_{k+1})}$$
$$= \sum_{\substack{\text{per} \\ y_1 < y_2 < \cdots < y_k}}' \frac{1}{\prod_1^k y_i}\left[\prod_{i>j}^k \frac{y_i - y_j}{y_i + y_j}\right]\frac{(-1)^{t(\text{per})}}{y_1 + y_2 + \cdots y_{k+1}},$$

where the sum on the right is over all permutations which have $y_1 < y_2 < y_3 < \cdots < y_k$. This means that the sum has only $k + 1$ terms; these terms arise from putting y_{k+1} equal to $g_1, g_2, \cdots, g_{k+1}$ in turn and arranging the other g's in ascending order. Thus the right side of this last relation may be written

$$\sum_{y=g_1}^{g_{k+1}} \frac{y}{\prod_1^{k+1} g_i}\left[\prod_{i>j}^{k+1} \frac{g_i - g_j}{g_i + g_j}\right]\left[\prod_{\substack{j=1 \\ g_j \neq y}}^{k+1} \frac{y + g_j}{|y - g_j|}\right]\frac{(-1)^{k+1-j}}{\sum_1^{k+1} g_i}$$
$$= \left[\left(\prod_{i>j}^{k+1} \frac{g_i - g_j}{g_i + g_j}\right)\Big/\prod_1^{k+1} g_i\right]\left[\sum_{i=1}^{k+1} g_i \left(\prod_{j \neq i} \frac{g_i + g_j}{g_i - g_j}\right)\frac{1}{\sum g_i}\right],$$

and the final bracket is unity in view of (b).

REFERENCES

[1] M. A. Girshick, "On the sampling theory of the roots of determinantal equations," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 203–224.
[2] R. A. Fisher, "The sampling distribution of some statistics obtained from non-linear equations," *Annals of Eugenics*, Vol. 9 (1939), pp. 238–249.
[3] P. L. Hsu, "On the distribution of roots of certain determinantal equations," *Annals of Eugenics*, Vol. 9 (1939), pp. 250–258.
[4] S. N. Roy, "p-statistics, or some generalizations on the analysis of variance appropriate to multivariate problems," *Sankhyā*, Vol. 3 (1939), pp. 341–396.
[5] H. Hotelling, "Relations between two sets of variates," *Biometrika*, Vol. 28 (1936), pp. 321–377.
[6] M. Bôcher, *Introduction to Higher Algebra*, Macmillan, 1929, p. 171.
[7] S. S. Wilks, "Certain generalizations in the analysis of variance," *Biometrika*, Vol. 24 (1932), pp. 471–494.

A BIVARIATE EXTENSION OF THE *U* STATISTIC[1]

By D. R. Whitney

Ohio State University

1. Summary. Let x, y, and z be three random variables with continuous cumulative distribution functions f, g, and h. In order to test the hypothesis $f = g = h$ under certain alternatives two statistics U, V based on ranks are proposed.

Recurrence relations are given for determining the probability of a given (U, V) in a sample of l x's, m y's, n z's and the different moments of the joint distribution of U and V. The means, second, and fourth moments of the joint distribution are given explicitly and the limit distribution is shown to be normal.

As an illustration the joint distribution of U, V is given for the case $(l, m, n) = (6, 3, 3)$ together with the values obtained by using the bivariate normal approximation. Tables of the joint cumulative distribution of U, V have been prepared for all cases where $l + m + n \leqq 15$.

2. Introduction. Let x, y, and z be three random variables with continuous cumulative distributions f, g, h. We wish to test the hypothesis that $f = g = h$ with the alternative that $f > g, f > h$, or say, $f > g > h$.

To test such a hypothesis with a sample of l x's, m y's, and n z's, we arrange the sample values in ascending order and let U count the number of times a y precedes an x, and V count the number of times a z precedes an x. As a critical region for the hypothesis with the alternative $f > g, f > h$ we propose to use $U \leqq K_1$, $V \leqq K_2$; or with the alternative $g > f > h$, $U \geqq K_3$, $V \leqq K_4$, where the constants K_i are chosen to give the correct significance level. Even if the significance level is fixed the constants K_i are not uniquely determined. A reasonable principle to follow in this case would be to choose

$$P(U \leqq K_2) \doteq P(V \leqq K_2) \quad \text{or} \quad P(U \geqq K_3) \doteq P(V \leqq K_4)$$

according to which alternative is chosen. In particular, if $m = n$ this leads to $K_1 = K_2$ and $K_3 + K_4 = m \cdot n$.

3. Moments of the joint distribution of *U* and *V*. We consider sequences of l x's, m y's, n z's and let $T_{lmn}(U, V)$ be the number of such sequences in which a y precedes an x U times and a z precedes an x V times. Omitting the last term in such a sequence leads to the relation

$$(1) \quad T_{lmn}(U, V) = T_{l-1,mn}(U - m, V - n) + T_{l,m-1,n}(U, V) + T_{lm,n-1}(U, V),$$

[1] The U statistic was introduced by H. B. Mann and the author in "On a test of whether one of two random variables is stochastically larger than the other," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 50–60. The present extension was carried out at the suggestion of J. W. Tukey, Princeton University.

where $T_{lmn}(U, V) = 0$ if $U < 0$; $V < 0$; $U > 0$, $m = 0$; or $V > 0$, $n = 0$; and $T_{0mn} = \binom{m+n}{m}$.

Under the hypothesis any of the $(l + m + n)!/l!m!n!$ sequences has equal probability. Hence

$$
(2) \qquad \begin{aligned} p_{lmn}(U, V) = {} & \frac{l}{l+m+n} p_{l-1,mn}(U - m, V - n) \\ & + \frac{m}{l+m+n} p_{l,m-1,n}(U, V) + \frac{n}{l+m+n} p_{lm,n-1}(U, V), \end{aligned}
$$

where $p_{lmn}(U, V)$ denotes the probability of a sequence of l x's, m y's, n z's having y precede an x U times and z precede an x V times.

To obtain the mean of U we multiply (2) by U and sum over all U, V. This gives

$$
(3) \qquad \begin{aligned} E_{lmn}(U) = {} & \frac{l}{l+m+n} E_{l-1,mn}(U) + \frac{m}{l+m+n} E_{l,m-1,n}(U) \\ & + \frac{n}{l+m+n} E_{lm,n-1}(U) + \frac{lm}{l+m+n}. \end{aligned}
$$

This and a similar equation for $E_{lmn}(V)$, together with the obvious initial conditions, give

$$
(4) \qquad E_{lmn}(U) = \frac{lm}{2}, \qquad E_{lmn}(V) = \frac{ln}{2}.
$$

The recurrence relations for the higher moments about the mean are obtained by multiplying (2) by $(U - \frac{1}{2}lm)^i(V - \frac{1}{2}ln)^j$ and summing over all U, V. Using $u = U - \frac{1}{2}lm$, $v = V - \frac{1}{2}ln$,

$$
(5) \qquad \begin{aligned} E_{lmn}(u^i v^j) = {} & \frac{l}{l+m+n} \sum_{\alpha=0}^{i} \sum_{\beta=0}^{j} \binom{i}{\alpha}\binom{j}{\beta}\left(\frac{m}{2}\right)^{i-\alpha}\left(\frac{n}{2}\right)^{j-\beta} E_{l-1,mn}(u^\alpha v^\beta) \\ & + \frac{m}{l+m+n} \sum_{\alpha=0}^{i} \binom{i}{\alpha} (-1)^{i-\alpha} \left(\frac{l}{2}\right)^{i-\alpha} E_{l,m-1,n}(u^\alpha v^j) \\ & + \frac{n}{l+m+n} \sum_{\beta=0}^{j} \binom{j}{\beta} (-1)^{j-\beta} \left(\frac{l}{2}\right)^{j-\beta} E_{lm,n-1}(u^i v^\beta). \end{aligned}
$$

For $i + j \leqq 4$ the solutions of (5) satisfying the initial conditions $E_{0mn}(u^i v^j) = E_{l00}(u^i v^j) = 0$ are found to be

$$
\begin{aligned} E_{lmn}(u^2) &= \frac{1}{12} lm(l + m + 1), \\ E_{lmn}(uv) &= \frac{1}{12} lmn, \\ E_{lmn}(u^2 v) &= E_{lmn}(uv^2) = 0, \end{aligned}
$$

$$
\begin{aligned}
E_{lmn}(u^4) &= \frac{1}{240}\,lm \\
&\qquad\cdot(l+m+1)(5l^2m+5lm^2-2l^2-2m^2+3lm-2l-2m), \\
E_{lmn}(u^3v) &= \frac{1}{240}\,lmn(5l^2m+5lm^2-2l^2-2m^2+3lm-2l-2m), \\
E_{lmn}(u^2v^2) &= \frac{1}{720}\,lmn \\
&\qquad\cdot(5l^3+5l^2m+5l^2n+15lmn+14l^2 \\
&\qquad\qquad+3lm+3ln-6mn+7l-2m-2n-2).
\end{aligned}
\tag{6}
$$

From symmetry considerations it follows that $E_{lmn}(u^iv^j) = 0$ if $i + j$ is odd.

4. Limit distribution of u and v. Let $F(l, m, n)$ be a function of integers l, m, n and define an operator Ψ by

$$
\begin{aligned}
\Psi F(l, m, n) \equiv{}& l[F(l, m, n) - F(l-1, m, n)] \\
&+ m[F(l, m, n) - F(l, m-1, n)] + n[F(l, m, n) - F(l, m, n-1)].
\end{aligned}
\tag{7}
$$

This permits (5) to be rewritten as

$$
\begin{aligned}
\Psi E_{lmn}(u^iv^j) ={}& l \sum_{\substack{\alpha=0 \\ \alpha+\beta<i+j}}^{i} \sum_{\beta=0}^{j} \binom{i}{\alpha}\binom{j}{\beta}\left(\frac{m}{2}\right)^{i-\alpha}\left(\frac{n}{2}\right)^{j-\beta} E_{l-1,mn}(u^\alpha v^\beta) \\
&+ m \sum_{\alpha=0}^{i-1} \binom{i}{\alpha}(-1)^{i-\alpha}\left(\frac{l}{2}\right)^{i-\alpha} E_{l,m-1,n}(u^\alpha v^j) \\
&+ n \sum_{\beta=0}^{j-1} \binom{j}{\beta}(-1)^{j-\beta}\left(\frac{l}{2}\right)^{j-\beta} E_{lm,n-1}(u^iv^\beta).
\end{aligned}
\tag{8}
$$

In order to work with equation (8) we need these properties:

(a) If $\Psi F(l, m, n)$ is a polynomial of degree t in all the variables, α in l, β in m, γ in n, then $F(l, m, n)$ is a polynomial of degree t in all the variables, α in l, β in m, γ in n.

(b) If P_t, Q_t are polynomials of degree t and $l, m, n \to \infty$ so that $F(l, m, n) \to F_0$ and $\dfrac{\Psi P_t}{Q_t} \to c$, then $\dfrac{P_t}{Q_t} \to \dfrac{c}{t}$.

Leaving the proof of these statements to a later section (Section 5), we shall apply them now to equation (8). Since $E_{lmn}(u^iv^j) = 0$ for $i + j$ odd we consider only the case $i + j = 2r$. For $r = 1, 2$, $E_{lmn}(u^iv^j)$ is a polynomial of degree $3r$, of degree at most $2r$ in l, i in m, and j in n. If we assume this to hold for $i + j < 2r$, then from (8) $\Psi E_{lmn}(u^iv^j)$, $i + j = 2r$, has these properties and hence $E_{lmn}(u^iv^j)$ does also.

In what follows there are two cases according as i and j are both even or both

odd. We give only the first case explicitly. Replacing i and j in (8) by $2i$ and $2j$, we obtain

$$\begin{aligned}\Psi E_{lmn}(u^{2i}v^{2j}) &= l\left[\binom{2i}{2i-2}\binom{m}{2}^2 E_{l-1,mn}(u^{2i-2}v^{2j})\right.\\ &+\binom{2i}{2i-1}\binom{2j}{2j-1}\binom{m}{2}\binom{n}{2} E_{l-1,mn}(u^{2i-1}v^{2j-1})\\ &\left.+\binom{2j}{2j-2}\binom{n}{2}^2 E_{l-1,mn}(u^{2i}v^{2j-2})\right]\\ &+m\left[\binom{2i}{2i-2}\binom{l}{2}^2 E_{l,m-1,n}(u^{2i-2}v^{2j})\right]\\ &+n\left[\binom{2j}{2j-2}\binom{l}{2}^2 E_{lm,n-1}(u^{2i}v^{2j-2})\right]+P_{3(i+j)-1}(l,m,n),\end{aligned}\tag{9}$$

where $P_{3(i+j)-1}(l, m, n)$ indicates a polynomial of degree $3(i + j) - 1$ in l, m, n, which is also of degree at most $2(i + j)$ in l, $2i$ in m, and $2j$ in n. This may be reduced to

$$\begin{aligned}\Psi E_{lmn}(u^{2i}v^{2j}) &= \tfrac{1}{4}lm(l+m)i(2i-1)E_{lmn}(u^{2i-2}v^{2j})\\ &+ lmn\cdot i\cdot jE_{lmn}(u^{2i-1}v^{2j-1}) + \tfrac{1}{4}ln(l+n)j(2j-1)E_{lmn}(u^{2i}v^{2j-2})\\ &+P_{3(i+j)-1}(l, m, n).\end{aligned}\tag{10}$$

Now we write
$$\lambda_{lmn}^{\alpha\beta} \equiv \frac{E_{lmn}(u^\alpha v^\beta)}{\sigma_u^\alpha \sigma_v^\beta};$$

then dividing (10) by $\sigma_u^{2i}\sigma_v^{2j} = [\frac{1}{12}\, lm(l+m+1)]^i\, [\frac{1}{12}\, ln(l+n+1)]^j$ gives

$$\begin{aligned}\frac{\Psi E_{lmn}(u^{2i}v^{2j})}{\sigma_u^{2i}\sigma_v^{2j}} &= \frac{\frac{1}{4}lm(l+m)i(2i-1)}{\frac{1}{12}lm(l+m+1)}\lambda_{lmn}^{2i-2,2j}\\ &+\frac{lmn\cdot i\cdot j}{\frac{1}{12}\sqrt{lm(l+m+1)ln(l+n+1)}}\lambda_{lmn}^{2i-1,2j-1}\\ &+\frac{\frac{1}{4}ln(l+n)j(2j-1)}{\frac{1}{12}ln(l+n+1)}\lambda_{lmn}^{2i,2j-2}+\frac{P_{3(i+j)-1}(l, m, n)}{\sigma_u^{2i}\sigma_v^{2j}}.\end{aligned}\tag{11}$$

Let
$$\rho(l, m, n) \equiv \frac{E_{lmn}(u, v)}{\sigma_u\sigma_v} = \sqrt{\frac{mn}{(l+m+1)(l+n+1)}}$$

and use $\rho(l, m, n) \to \rho_0$ to mean $l, m, n \to \infty$ in such a way that
$$\sqrt{\frac{mn}{(l+m+1)(l+n+1)}} \to \rho_0.$$

We then have for $\rho(l, m, n) \to \rho_0$

$$\lambda_{lmn}^{11} \to \rho_0,\ \lambda_{l,m,n}^{40} \to 3,\ \lambda_{lmn}^{31} \to 3\rho_0,\ \lambda_{lmn}^{22} \to 1 + 2\rho_0^2.\tag{12}$$

Dropping the l, m, n to denote the limiting values, (12) are just special cases i.e., $i + j = 2$ or 4, of

$$\lambda^{2i,2j} = \frac{(2i)!(2j)!}{2^{i+j}} \sum_{\alpha=0}^{\min(i,j)} \frac{(2\rho_0)^{2\alpha}}{(i-\alpha)!(j-\alpha)!(2\alpha)!},$$

(13)

$$\lambda^{2i+1,2j+1} = \frac{(2i+1)!(2j+1)!}{2^{i+j}} \rho_0 \sum_{\alpha=0}^{\min(i,j)} \frac{(2\rho_0)^{2\alpha}}{(i-\alpha)!(j-\alpha)!(2\alpha+1)!}.$$

Inductively then, we assume for $\alpha + \beta < 2(i + j)$ that $\lambda_{lmn}^{\alpha\beta}$ satisfies (13) for $\rho(l, m, n) \to \rho_0$. Since $P_{3(i+j)-1}(l, m, n)$ in (11) has degree at most $2(i + j)$ in l, $2i$ in m, $2j$ in n, we obtain

$$\operatorname*{Lim}_{\rho(l,m,n)\to\rho_0} \frac{\Psi E_{lmn}(u^{2i}v^{2j})}{\sigma_u^{2i}\sigma_v^{2j}}$$

(14)

$$= 3i(2i-1)\frac{(2i-2)!(2j)!}{2^{i+j-1}} \sum_{\alpha=0}^{\min(i-1,j)} \frac{(2\rho_0)^{2\alpha}}{(i-1-\alpha)!(j-\alpha)!(2\alpha)!}$$

$$+ 12i\cdot j\rho_0 \frac{(2i-1)!(2j-1)!}{2^{i+j-2}} \rho_0 \sum_{\alpha=0}^{\min(i-1,j-1)} \frac{(2\rho_0)^{2\alpha}}{(i-1-\alpha)!(j-1-\alpha)!(2\alpha+1)!}$$

$$+ 3j(2j-1)\frac{(2i)!(2j-2)!}{2^{i+j-1}} \sum_{\alpha=0}^{\min(i,j-1)} \frac{(2\rho_0)^{2\alpha}}{(i-\alpha)!(j-1-\alpha)!(2\alpha)!},$$

and this reduces to

$$(15) \qquad \operatorname*{Lim}_{\rho(l,m,n)\to\rho_0} \frac{\Psi E_{lmn}(u^{2i}v^{2j})}{\sigma_u^{2i}\sigma_v^{2i}} = 3(i+j)\frac{(2i)!(2j)!}{2^{i+j}} \sum_{\alpha=0}^{\min(i,j)} \frac{(2\rho_0)^{2\alpha}}{(i-\alpha)!(j-\alpha)!(2\alpha)!}.$$

From this it follows that

$$(16) \qquad \operatorname*{Lim}_{\rho(l,m,n)\to\rho_0} \frac{E_{lmn}(u^{2i}v^{2j})}{\sigma_u^{2i}\sigma_v^{2j}} = \frac{(2i)!(2j)!}{2^{i+j}} \sum_{\alpha=0}^{\min(i,j)} \frac{(2\rho_0)^{2\alpha}}{(i-\alpha)!(j-\alpha)!(2\alpha)!},$$

and in like manner for $E_{lmn}(u^{2i+1}v^{2j+1})$. Therefore the moments of the limit distribution are those of a bivariate normal distribution. Hence the variables

$$\frac{U - \frac{lm}{2}}{\sqrt{\frac{1}{12}lm(l+m+1)}}, \qquad \frac{V - \frac{ln}{2}}{\sqrt{\frac{1}{12}ln(l+n+1)}}$$

have in the limit a joint normal distribution with means 0, variances 1, and correlation coefficient ρ_0, where

$$\rho_0 = \operatorname*{Lim}_{l,m,n\to\infty} \sqrt{\frac{mn}{(l+m+1)(l+n+1)}}.$$

5. Properties of Ψ.

LEMMA 1. *If*

$$F(x, y, z) = \sum_{i=0}^{\lambda}\sum_{j=0}^{\mu}\sum_{k=0}^{\nu} a_{ijk}\, x^i y^j z^k,$$

then

$$\Psi F(x, y, z) = \sum_{i=0}^{\lambda} \sum_{j=0}^{\mu} \sum_{k=0}^{\nu} A_{ijk} x^i y^j z^k,$$

where

$$A_{ijk} = \sum_{\alpha=i}^{\lambda} (-1)^{\alpha-i} a_{\alpha jk} \binom{\alpha}{i-1} + \sum_{\beta=j}^{\mu} (-1)^{\beta-j} a_{i\beta k} \binom{\beta}{j-1} + \sum_{\gamma=k}^{\nu} (-1)^{\gamma-k} a_{ij\gamma} \binom{\gamma}{k-1}.$$

This follows from a straightforward application of the definition of Ψ.

LEMMA 2. *If $F(x, y, z)$ is a polynomial in x, y, z of degree σ, of degree λ in x, μ in y, ν in z, then so is ΨF.*

This follows from the representation of ΨF in Lemma 1.

LEMMA 3. *For any polynomial $F(x, y, z)$ there exists a polynomial $G(x, y, z)$ such that $\Psi G = F$.*

Let the coefficients of F be denoted by A_{ijk} and the unknown coefficients of G by a_{ijk}. The lemma will follow if we can solve the equations in Lemma 1 for a_{ijk}, $i = 0, 1, \cdots, \lambda; j = 0, 1, \cdots, \mu; k = 0, 1, \cdots, \nu$. For $i + j + k$ a maximum for all the i, j, k of A_{ijk}, we have

$$A_{ijk} = a_{ijk} \binom{i}{i-1} + a_{ijk} \binom{j}{j-1} + a_{ijk} \binom{k}{k-1} = (i + j + k)a_{ijk}.$$

By induction we assume that the equation can be solved for the a_{ijk} for all i, j, k such that $i + j + k > t$. Then for $i + j + k = t$ we have $A_{ijk} = (i + j + k)a_{ijk}$ plus a's whose subscripts add to more than t. Hence the a_{ijk} can be determined.

LEMMA 4. *If $\Psi[F(x, y, z) - G(x, y, z)] = 0$, then $F(x, y, z) - F(0, 0, 0) = G(x, y, z) - G(0, 0, 0)$.*

Let $t = x + y + z$. The lemma is true for $t = 0$, and we assume it to be true for all x, y, z such that $x + y + z < t$. Then for $x + y + z = t$,

$$\Psi[F(x, y, z) - G(x, y, z)] = 0$$

gives

$$(x + y + z)[F(x, y, z) - G(x, y, z)] - x[F(x - 1, y, z) - G(x - 1, y, z)] - y[F(x, y - 1, z) - G(x, y - 1, z)] - Z[F(x, y, z - 1) - G(x, y, z - 1)] = 0.$$

Using our induction assumption,

$$(x + y + z)[F(x, y, z) - G(x, y, z)] = (x + y + z)[F(0, 0, 0) - G(0, 0, 0)],$$

and the lemma follows.

6. Distribution of u and v in a particular case with $l = 6, m = n = 3$ Using. the relation (1), the table of $T_{633}(U, V)$ (Table 1) was obtained. In this case $E(U) = E(V) = 9$, $\sigma_u^2 = \sigma_v^2 = 15$, $\sigma_{uv} = 4.5$, $\rho = 0.3$.

TABLE 1

$T_{633}(U, V)$

V \ U	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
9	14	15	32	55	78	103	150	155	178	200	178	155	150	103	78	55	32	15	14
8	17	18	41	65	91	112	158	160	194	178	173	144	139	95	71	46	30	14	14
7	16	20	42	66	85	108	146	158	160	155	144	122	110	74	52	38	23	11	10
6	25	24	48	71	95	114	170	146	158	150	139	110	103	64	49	33	21	10	10
5	20	20	39	58	75	98	114	108	112	103	95	74	64	43	31	22	13	6	5
4	19	18	37	51	74	75	95	85	91	78	71	52	49	31	23	14	9	4	4
3	15	16	32	56	51	58	71	66	65	55	46	38	33	22	14	10	6	3	3
2	14	15	30	32	37	39	48	42	41	32	30	23	21	13	9	6	4	2	2
1	10	12	15	16	18	20	24	20	18	15	14	11	10	6	4	3	2	1	1
0	20	10	14	15	19	20	25	16	17	14	14	10	10	5	4	3	2	1	1

TABLE 2

$$\sum_{V=0}^{k} \sum_{U=0}^{h} p_{633}(U, V)$$

k \ h	0	1	2	3	4	5	6	7	8	9
9	9 (12)	18 (21)	36 (37)	62 (59)	96 (91)	137 (132)	191 (181)	242 (237)	298 (295)	351 (352)
8	8 (10)	17 (19)	33 (33)	56 (52)	86 (80)	120 (114)	166 (156)	210 (201)	256 (249)	298 (295)
7	7 (9)	15 (16)	29 (28)	48 (45)	73 (67)	102 (95)	139 (128)	174 (165)	210 (201)	242 (237)
6	7 (8)	13 (14)	24 (23)	41 (36)	61 (54)	84 (76)	113 (101)	139 (128)	166 (156)	191 (181)
5	5 (6)	10 (11)	19 (18)	32 (28)	46 (41)	63 (57)	84 (76)	102 (95)	120 (114)	137 (132)
4	4 (5)	8 (8)	15 (13)	24 (21)	35 (30)	46 (41)	61 (54)	73 (67)	86 (80)	96 (91)
3	3 (3)	6 (6)	11 (10)	17 (14)	24 (21)	32 (28)	41 (36)	48 (45)	56 (52)	62 (59)
2	2 (2)	4 (4)	8 (6)	11 (10)	15 (13)	19 (18)	24 (23)	29 (28)	33 (33)	36 (37)
1	2 (1)	3 (3)	4 (4)	6 (6)	8 (8)	10 (11)	13 (14)	15 (16)	17 (19)	18 (21)
0	1 (1)	2 (1)	2 (2)	3 (3)	4 (5)	5 (6)	7 (8)	7 (9)	8 (10)	9 (12)

Table 2 gives the cumulative distribution $\sum_{V=0}^{k} \sum_{U=0}^{h} p_{633}(U, V)$. The numbers have all been multiplied by 1000. The figures in parentheses are the values obtained by using $(U - 9 - \frac{1}{2})/\sqrt{15}$, $(V - 9 - \frac{1}{2})/\sqrt{15}$ as random variables from a bivariate normal distribution of means zero, variances one, and correlation coefficient 0.3.

7. Example. Suppose that y, x, z denote the lengths of life of rats that have been exposed to insecticides of supposedly decreasing toxicity. We would then be interested in the hypothesis $g = f = h$ under the alternative $g > f > h$.

For a sample of 3 y's, 6 x's, and 3 z's, a critical region of size .044 is found from the preceding table to be $U \geqq 12$, $V \leqq 6$. In an experiment the sequence

TABLE 2—*Continued*

$$\sum_{V=0}^{k} \sum_{U=0}^{h} p_{633}(U, V)$$

k \ h	10	11	12	13	14	15	16	17	18
9	400 (404)	440 (447)	478 (482)	502 (508)	520 (526)	533 (537)	541 (544)	544 (548)	548 (551)
8	338 (336)	369 (370)	398 (398)	418 (417)	431 (430)	441 (439)	447 (444)	450 (446)	452 (449)
7	272 (268)	296 (293)	318 (313)	332 (327)	342 (337)	349 (345)	353 (346)	355 (348)	357 (349)
6	213 (204)	230 (222)	246 (236)	256 (245)	263 (252)	268 (255)	271 (258)	272 (259)	274 (260)
5	151 (147)	162 (159)	173 (168)	179 (174)	184 (178)	187 (180)	189 (182)	190 (183)	190 (183)
4	106 (101)	113 (108)	119 (114)	124 (118)	127 (120)	129 (121)	130 (122)	130 (123)	131 (123)
3	68 (65)	72 (70)	76 (73)	79 (75)	81 (76)	82 (77)	83 (78)	83 (78)	83 (78)
2	39 (40)	42 (43)	44 (44)	45 (45)	46 (46)	47 (47)	47 (47)	47 (47)	48 (47)
1	20 (23)	21 (24)	22 (25)	23 (26)	23 (26)	23 (26)	24 (26)	24 (27)	24 (27)
0	10 (13)	10 (13)	11 (13)	11 (14)	12 (14)	12 (14)	12 (14)	12 (14)	12 (14)

$yyxxxyxzxzzzx$ was obtained. For this sample $U = 15$, $V = 4$, and consequently we presume the toxic effects to be as supposed.

For a sample of 7 y's, 6 x's, 8 z's, we first compute

$$E(U) = 21, \qquad E(V) = 28, \qquad \sigma_U^2 = 49, \qquad \sigma_V^2 = 60, \qquad \rho^2 = 4/15.$$

The critical region can be written as

$$\frac{U - E(U)}{\sigma_U} \geqq h, \qquad \frac{V - E(V)}{\sigma_V} \leqq -k,$$

where h, k are to be determined to give a significance level of 5%, say, and subject to

$$P\left(\frac{U - E(U)}{\sigma_U} \geqq h\right) \doteq P\left(\frac{V - E(V)}{\sigma_V} \leqq -k\right).$$

With the normal approximation to the distribution of U or V the last condition implies $h = k$. Then entering Pearson's table for the normal bivariate distribution with $\rho = -.52$ (interpolating between $-.50$ and $-.55$) we find that $h = k \doteq .37$ are the desired values. This gives a 5% critical region of $U \geqq 24$, $V \leqq 25$.

RELATIONS BETWEEN VARIOUSLY DEFINED EFFECTS AND INTERACTIONS IN ANALYSIS OF VARIANCE

By S. Vajda

London, England

1. Summary. From an algebraic point of view the analysis of variance tests of effects and interactions can be based on the minimum values of a certain quadratic expression in which the "h-matrix" (defined in Section 3) is fundamental. The arbitrariness in the choice of this matrix reflects the arbitrariness in the definition of effects and interactions. The paper considers the dependence of the result of these tests on the h-matrix used and expresses the answer by the two theorems of Section 4, which are proved in the subsequent sections.

2. Introduction. The sums of squares which appear in an analysis of variance when the significance of effects and/or interactions is tested can be obtained by taking the minima with regard to values $a_{k_1 \cdots k_s}$ of such expressions as

$$(1) \qquad \sum_{i_1=1}^{n_1} \cdots \sum_{i_s=1}^{n_s} g_{i_1 \cdots i_s} [y_{i_1 \cdots i_s} - \sum_{k_1} \cdots \sum_{k_s} a_{k_1 \cdots k_s} h_{k_1 \cdots k_s}(i_1 \cdots i_s)]^2,$$

where the $y_{i_1 \cdots i_s}$ are the means of $g_{i_1 \cdots i_s}$ observed values for levels i_t of variables x_t $(t = 1, \cdots, s)$, respectively, and the values h form a nonsingular matrix which will be described in detail in the next section. The summation inside the bracket in (1) is carried out over sets $(k_1 \cdots k_s)$ of subscripts, $0 \leq k_t \leq n_t - 1$, which depend on the aggregate of effects and interactions to be tested. If all $n_1 \cdots n_s$ possible sets appeared in (1), the minimum would, of course, be zero. To each test there belongs a set of k's which is left out of the combinations of subscripts in (1) according to the following rule:

The interaction of order $(t - 1)$ between $x_1, \cdots, x_t$ is tested by omitting all $a_{k_1 \cdots k_t 0 \cdots 0}$ for which $k_1 \cdots k_t \neq 0$. (A main effect is equivalent to an interaction of order zero.) An aggregate of interactions is tested by leaving out all combinations referring to any of its several components [2].

As an illustration, let us take $s = 2$ and $g_{i_1 \cdots i_s} = 1$. We choose the following (orthogonal) matrix of $h_{k_1 k_2}(i_1 i_2)$:

$$\begin{array}{rl} (i_1 i_2) = & \begin{matrix} 11 & 12 & 21 & 22 & 31 & 32 \end{matrix} \\ \begin{matrix} (k_1 k_2) = 00 \\ 01 \\ 10 \\ 11 \\ 20 \\ 21 \end{matrix} & \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & -2 & -2 \\ 1 & -1 & 1 & -1 & -2 & 2 \\ 1 & 1 & -1 & -1 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 \end{pmatrix}. \end{array}$$

If we test for the first main effect, then we retain a_{00}, a_{01}, a_{11}, and a_{21}, and we obtain for the minimum, after straightforward calculations,

$$\frac{(y_{11} + y_{12})^2 + (y_{21} + y_{22})^2 + (y_{31} + y_{32})^2}{2} - \frac{(\Sigma\Sigma\, y_{i_1 i_2})^2}{6}.$$

Similarly, testing for the interaction, we would retain a_{00}, a_{01}, a_{10}, a_{20}, and obtain

$$\Sigma\Sigma\, y^2_{i_1 i_2} - \tfrac{1}{2}[(y_{11} + y_{12})^2 + (y_{21} + y_{22})^2 + (y_{31} + y_{32})^2]$$
$$-\tfrac{1}{3}[(y_{11} + y_{21} + y_{31})^2 + (y_{12} + y_{22} + (y_{32})^2] + (\Sigma\Sigma\, y_{i_1 i_2})^2/6.$$

If we had taken general weights $g_{i_1 \cdots i_s}$, but still using the same values for the h-matrix, then we should have obtained results which are equivalent to those given by Yates' "method of weighted squares of means" [5].

3. Assumptions and definitions. Let the "h-matrix" $h_{k_1 \cdots k_s}(i_1 \cdots i_s)$ $(i_t = 1, \cdots, n_t;\, k_t = 0, 1, \cdots, n_t - 1)$ be such that all the elements in the same row have equal sets of subscripts and all the elements in the same column have equal sets of arguments. It will be assumed that this matrix satisfies the following conditions:

CONDITION A.

$$\sum_{i_1=1}^{n_1} \cdots \sum_{i_s=1}^{n_s} w_{i_1 \cdots i_s} h_{k_1 \cdots k_s}(i_1 \cdots i_s)\, h_{m_1 \cdots m_s}(i_1 \cdots i_s) \neq 0$$

if simultaneously $k_t = m_t$, and $= 0$ otherwise. The $w_{i_1 \cdots i_s}$ are positive weights. It follows that the h-matrix is not singular.

CONDITION B. *If any $k_t = 0$, then $h_{k_1 \cdots k_t \cdots k_s}$ is independent of i_t.*

In particular, if the h's are orthogonal polynomials of degrees k_t in i_t, then Conditions A and B hold by definition.

It has been shown [1] that these two conditions can be satisfied simultaneously only if the weights are "proportionate," i.e., if $w_{i_1 \cdots i_s}/\sum_{i_t=1}^{n_t} w_{i_1 \cdots i_s}$ is independent of all i_m $(m \neq t)$ for all t.

From Condition A can be derived the following lemma, which will be used at a later state:

LEMMA. *If $k_t \neq 0$, then $\sum_{i_t=1}^{n_t} w_{i_1 \cdots i_s}\, h_{k_1 \cdots k_s}(i_1 \cdots i_s) = 0$.*

PROOF. We assume that $t = 1$; this clearly does not restrict the generality of the argument, since it may be repeated identically for any other value of t. From Condition A we have the equations:

$$\sum_{i_1=1}^{n_1} \cdots \sum_{i_s=1}^{n_s} w_{i_1 \cdots i_s} h_{k_1 \cdots k_s}(i_1 \cdots i_t)\, h_{0m_2 \cdots m_s}(i_1 \cdots i_s) = 0$$

for all $m_2, \cdots, m_s$, since k_1 is assumed to be different from zero. If we regard the $n_2 \cdots n_s$ expressions $\sum_{i_1=1}^{n_t} w_{i_1 \cdots i_s}\, h_{k_1 \cdots k_s}(i_1 \cdots i_s)$ for all $i_2, \cdots, i_s$ as unknown values, we have the same number of linear homogeneous equations for them. The determinant of the system is orthogonal and hence not zero. It follows that the unknown values must be zero, and thus the lemma is proved.

All sets $(k_1 \cdots k_t 0 \cdots 0)$ with $k_1 \cdots k_t \neq 0$ form a "block," which we denote by $((k_1 \cdots k_t))$. The meaning of $((k_{m_1} \cdots k_{m_t}))$ is immediately obvious. Every set of subscripts belongs to one and only one block. If we consider a particular block and then omit one or more values from within the double brackets denoting it, another block is obtained, which we call a "sub-block" of the former.

4. The problem. Even with Conditions A and B to be satisfied, there remains still an arbitrariness in the choice of the h-matrix, and this reflects an arbitrariness in the definition of interactions [3]. If the h's are such that Condition A is satisfied with $w_{i_1 \cdots i_s} = g_{i_1 \cdots i_s}$ for all i_t, then the computation of the minimum of (1) becomes very simple, but clearly this cannot be a reason for choosing the h-matrix accordingly [4]. However, in this paper we shall be concerned with another aspect of the situation: we wish to find out whether two different h-matrices can lead to the same minimum value and, if so, under what conditions. The answer depends on the particular test carried out and is expressed by the following two theorems.

THEOREM 1. *If two h-matrices satisfy Condition A with regard to the same weights, then both lead to the same minimum of* (1), *whatever the aggregate of interactions tested.*

THEOREM 2. *If the aggregate is such that for each retained block of subscripts all its sub-blocks are also retained, then the minimum of* (1) *is independent of the h-matrix* (*even if the latter is not orthogonal with regard to any weights*).

It follows from the latter theorem that when only the highest order interaction (that of order $s - 1$) is tested, and hence all sets of subscripts except those constituting the block $((k_1 \cdots k_s))$ are retained, any h-matrix leads to the same conclusion.

5. Transformation of the problem. In what follows we shall denote, where no misunderstanding can arise, the various sets $(i_1 \cdots i_s)$ by $I_1, \cdots, I_N$, and the sets $(k_1 \cdots k_s)$ by $K_1, \cdots, K_M$. Here $N = n_1 \cdots n_s$, and M is the number of retained sets of subscripts, e.g., $M = (n_1 - 1) \cdots (n_t - 1)$ if only the block $((k_1 \cdots k_t))$ were retained. We have $N > M$, except in the trivial case where the minimum of (1) is zero.

Let us now imagine that we have two h-matrices, the elements of which are denoted by h and h' respectively. If for any given set of $a_{K_i}(i = 1, \cdots, M)$ we can find a set a'_{K_i} so that

$$\sum_{i=1}^{M} a_{K_i} h_{K_i}(I_t) = \sum_{i=1}^{M} a'_{K_i} h'_{K_i}(I_t) \tag{2}$$

for all $I_t(t = 1, \cdots, N)$, then clearly the set of values which (1) can assume is identical with that of a similar expression when h is replaced by h'. Hence the minima of the two expressions will also be the same.

It follows that different h-matrices will lead to the same minimum of (1) if they and the retained blocks of subscripts are such that (2) can be solved for the a'_{K_i}, assuming that the a_{K_i} are given. In (2) there are N equations for the M

unknowns a'_{K_i}. It will be possible, therefore, to solve the set only if not more than M of the equations are linearly independent.

Regarding, to begin with, only the left-hand side (l.h.s.) of (2), we can certainly select M sets of arguments $I_1, \cdots, I_M$ so that the determinant $| h_{K_i}(I_t) | \neq 0$ (since the complete h-matrix is not singular). Hence for any further argument J, say, we can solve the system of linear equations

$$\begin{aligned} h_{K_1}(J) &= \sum_{t=1}^{M} C_{I_t} h_{K_1}(I_t), \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ h_{K_M}(J) &= \sum_{t=1}^{M} C_{I_t} h_{K_M}(I_t). \end{aligned} \tag{3}$$

This gives $h_{K_i}(J)$ as a linear combination of $h_{K_i}(I_1), \cdots, h_{K_i}(I_M)$ which is the same for all $i = 1, \cdots, M$. Therefore the l.h.s. of those equations in (2) in which $I_t = J$ will again be the same linear combination of the l.h.s. of the equations in which the arguments are $I_1, \cdots, I_M$, respectively. Consequently the whole equation, written for J, will be the very same linear combination of the equations for the I_t severally, if it can be shown that the C_{I_t} which we find from (3) are equally applicable to the r.h.s. of (2), i.e., to the h'_{K_i}. Since the two matrices are, by the assumptions of Theorem 1, orthogonal with regard to the same weights, it is sufficient to prove that the C_{I_t}, which are the solutions of (3), although possibly dependent on the weights $w_{i_1 \cdots i_s}$, do not otherwise depend on the h-matrix considered.

6. Proof of Theorem 1. In general, the sets of subscripts $K_1, \cdots, K_M$ will not all belong to the same block. We consider first the block out of which K_1 is taken and assume that it consists of $K_1, \cdots, K_P$ $(P \leq M)$. It is no restriction of generality to assume further that this is the block $((k_1 \cdots k_m))$, so that $P = (n_1 - 1) \cdots (n_m - 1)$.

We fix our attention on one single set belonging to this block, say $(\bar{k}_1, \cdots, \bar{k}_m, 0, \cdots, 0)$. Conditions A and B imply linear relations between the $h_{\bar{k}_1 \cdots \bar{k}_m 0 \cdots 0}(i_1 \cdots i_s)$, and we shall now establish how many of these values can be chosen independently, thereby fixing all others implicitly. If $h_{\bar{k}_1 \cdots \bar{k}_m 0 \cdots 0}$ is known for $(\bar{\imath}_1, \cdots, \bar{\imath}_s)$ where the $\bar{\imath}_t$ are some fixed values, then, by virtue of Condition B, it is also known for all $(\bar{\imath}_1, \cdots, \bar{\imath}_m, i_{m+1}, \cdots, i_s)$ where the $i_{m+1}, \cdots, i_s$ are arbitrary. We need therefore only investigate relations between the $n_1 \cdots n_m$ values $h_{\bar{k}_1 \cdots \bar{k}_m 0 \cdots 0}(i_1 \cdots i_m \bar{\imath}_{m+1} \cdots \bar{\imath}_s)$. These are not all independent either, since our lemma gives, for $r = 1, \cdots, m$,

$$\sum_{i_r=1}^{n_r} w_{i_1 \cdots i_s} h_{\bar{k}_1 \cdots \bar{k}_m 0 \cdots 0}(i_1 \cdots i_s) = 0 \tag{4}$$

for all $i_1, \cdots, i_{r-1}, i_{r+1}, \cdots, i_s$.

Thus only $(n_1 - 1) \cdots (n_m - 1) = P$ values among the $h_{\bar{k}_1 \cdots \bar{k}_m 0 \cdots 0}(i_1 \cdots i_s)$ will be independent, and it is easy to indicate how such a set can be found. In

the matrix $\| h_{K_j}(I_t) \|$ ($j = 1, \cdots, P$; $t = 1, \cdots, M$) there must be a square sub-matrix of order P which is not singular, since otherwise the determinant of (3) would be zero, contrary to our assumptions. Let this sub-matrix be $\| h_{K_j}(I_u) \|$ ($u = t_1, \cdots, t_P$; j as before). Then it follows that the $h_{k_1 \cdots k_m 0 \cdots 0}(I_u)$ constitute a set of values which can arbitrarily be selected. Indeed, if they were dependent, by virtue of (4) and Condition B, then identical linear relations would hold for all K_j, i.e., for all rows of the matrix $\| h_{K_j}(I_u) \|$, which would hence be singular.

We may, then, rewrite the first P equations in (3) by expressing all $h(I_t)$ on the r.h.s. in terms of $I_{t_1}, \cdots, I_{t_P}$ as arguments. The coefficients will be linear combinations of C_{I_t} and of the weights, which appear in (4). Since the l.h.s., i.e., $h(J)$, with subscripts of h as before, can also be expressed as a linear combination of these same $h(I_u)$, by virtue of (4) and Condition B, we see that the C_{I_t} must satisfy P identities, in which only the weights $w_{i_1 \cdots i_s}$ are parameters.

All blocks to which the $K_1, \cdots, K_M$ belong can be treated in the same way and thus we obtain altogether M equations from which the $C_{I_t}(t = 1, \cdots, M)$ can be obtained. They will depend on the weights, but not otherwise on the h-matrix. This completes the proof of Theorem 1.

7. Proof of Theorem 2. We turn now to Theorem 2, and assume that the sets of subscripts retained in (1) are those of the blocks $B_0, B_1, \cdots, B_n$ and of all their sub-blocks. We can at once indicate the sets of arguments $I_1, I_2, \cdots$ equal in number to the retained sets of subscripts, and such that $h(j_1 \cdots j_s)$ can be expressed as a linear combination of $h(I_1), h(I_2), \cdots$ for all sets of subscripts considered. For this purpose we take, for each retained set $(k_1 \cdots k_s)$, the set of arguments $(k_1 + 1, \cdots, k_s + 1)$. Thus there will be the same number ($=M$) of sets of arguments as there are sets of retained subscripts. We note in particular, that if any of the $k_t = 0$, the corresponding j_t will be unity. This will be the case in respect of all sets of a block, if it is true for any set in it.

To simplify our formulae, we introduce the following notation: If $(J) \equiv (j_1 \cdots j_s)$, then $(J)_i$ is the result of replacing by unity all those j_t which correspond to a $k_t = 0$ in block B_i. Further, $(J)_{ij}$ is the result of replacing by unity all those j_t which correspond to a $k_t = 0$ either in B_i or in B_j, or, in other words, those j_t which do not correspond to the largest common sub-block of B_i and B_j. The notation $(J)_{ij \cdots k}$ is similarly defined. Now if K_i is any set of subscripts in the block B_i, then it follows from Condition B that

$$h_{K_i}(j_1 \cdots j_s) = h_{K_i}(j_1 \cdots j_s)_i ,$$

and, more generally,

$$h_{K_i}(j_1 \cdots j_s)_{j \cdots k} = h_{K_i}(j_1 \cdots j_s)_{ij \cdots k} ,$$

since all those additional arguments 1 in $(j_1 \cdots j_s)_{ij \cdots k}$ which do not already appear in $(j_1 \cdots j_s)_{j \cdots k}$ correspond to zeros in the subscripts of B_i. Moreover, these relations remain true if we take, instead of B_i, any of its sub-blocks, since such a sub-block contains all those zeros which were in B_i (and some more).

We shall now prove that the relation

$$(5) \qquad h_K(J) = \sum_{i=0}^{n} h_K(J)_i - \sum_{i \neq j=0}^{n} h_K(J)_{ij} + \cdots + (-1)^n h_K(J)_{01\cdots n},$$

(J) being an arbitrary set of arguments, holds for all K out of $B_0, B_1, \cdots, B_n$ and also out of any of their sub-blocks. This is a linear relation of the type which we need for the proof of Theorem 2 and we see that all sets of arguments appearing on the r.h.s. are among those $I_1, I_2, \cdots$ which we have initially selected as a basis. Hence, if we prove that this relation holds for any K out of the blocks and sub-blocks considered, then Theorem 2 follows.

First let K be a set out of B_0. Equation (5) can be written as follows:

$$h_K(J) = h_K(J)_0 + \sum_{i=1}^{n} h_K(J)_i - \sum_{i=1}^{n} h_K(J)_{0i} - \sum_{i \neq j=1}^{n} h_K(J)_{ij}$$
$$+ \sum_{i \neq j=1}^{n} h_K(J)_{0ij} + \cdots + (-1)^n h_K(J)_{01\cdots n}.$$

Now we have $h_K(J) = h_K(J)_0$. Moreover, the second term on the r.h.s. cancels with the third, the fourth with the fifth, and so on until all terms are exhausted. This proves relation (5) for K out of B_0. But it is evident that the proof could equally well have been carried out for any other of the given blocks or for any of their sub-blocks. This completes the proof of Theorem 2. It will be noticed that no weights appear in (5), so that under the given conditions the theorem holds even for matrices which are not orthogonal (in the sense of Condition A) with regard to any set of weights.

REFERENCES

[1] S. VAJDA, "A note on the use of weighted orthogonal functions in statistical analysis," *Proc. Cambridge Philos. Soc.*, Vol. 44 (1948), p. 588.

[2] S. VAJDA, "An outline of the theory of the analysis of variance," *Jour. Inst. Actuaries Students' Soc.*, Vol. 7 (1948), p. 235.

[3] D. J. FINNEY, "Main effects and interactions," *Jour. Am. Stat. Assn.*, Vol. 43 (1948), p. 566.

[4] S. VAJDA, "Technique of the analysis of variance," *Nature*, Vol. 160 (1947), p. 27.

[5] F. YATES, "The analysis of multiple classifications with unequal numbers in the different classes," *Jour. Am. Stat. Assn.*, Vol. 29 (1934), p. 51.

ON UNIFORMLY CONSISTENT TESTS

By Agnes Berger

New York City

1. Introduction. If we wish to decide on the true distribution of a random variable known to be distributed according to one or the other of two given distributions F_0 and F_1, then, no matter how small a bound is given in advance, it is always possible to devise a test based on a sufficiently large number of independent observations for which the probabilities of erroneous decisions are smaller than the previously assigned bound. A sequence of tests for which the corresponding probabilities of errors tend to zero has been called consistent [1].[1]

Let us suppose now that all we know about the true distribution of some random variable is that it belongs to one of two given families of distributions and it is desired to decide which of the two it belongs to; i.e., we have to test a composite hypothesis. It may again be possible to construct a sequence of tests $\{T_j\}$, $j = 1, 2, \cdots$, such that for any $\epsilon > 0$ there exists an index N such that for $j > N$ the probabilities of errors corresponding to T_j are smaller than ϵ. The sequence $\{T_j\}$ may then be called uniformly consistent. Conditions under which uniformly consistent tests exist have been given by von Mises [5], and by Wald [2], [3], [4], as implied, for example, by his proof of the uniform consistency of the likelihood ratio test. In this paper a different set of conditions is given which do not restrict in any way the nature of the distribution functions considered. It is also shown that the conditions to be described are satisfied in a large class of cases occurring in practical statistics.

Since the results we are to prove have their counterpart in abstract measure theory we shall take advantage of that method. The reader will have no difficulty in establishing the correspondence between the statistical and measure theoretical formulation.

Notations. Let X be an arbitrary set and $\mathfrak{B}$ a Borel field of subsets B of X. Let $\mathfrak{M}(\mathfrak{B})$ be the family of all probability measures $m(B)$ defined on $\mathfrak{B}$, i.e., the family of all countably additive nonnegative set functions defined on $\mathfrak{B}$ for which $m(X) = 1$. Hereafter a "measure" will denote an element of $\mathfrak{M}(\mathfrak{B})$ and a "set of measures" a subset of $\mathfrak{M}(\mathfrak{B})$. For any positive integer k, let X^k be the kth direct product of X by itself, $\mathfrak{B}^k$ the kth direct product of $\mathfrak{B}$ by itself, $\mathfrak{C}^{\cdot k}$ the field consisting of all finite sums of sets of $\mathfrak{B}^k$ and $\mathfrak{B}^{\cdot k}$ the smallest σ-field containing $\mathfrak{C}^{\cdot k}$. For any measure m on $\mathfrak{B}$, we define m^k in the usual way as the unique measure defined on $\mathfrak{B}^{\cdot k}$ for which

$$m^k\left(\sum_{i=1}^{l} B_{i1}\cdot B_{i2}\cdot \cdots \cdot B_{ik}\right) = \sum_{i=1}^{l} m(B_{i1})m(B_{i2}) \cdots m(B_{ik})$$

for any disjoint system $B_{ij} \varepsilon \mathfrak{B}$, $i = 1, 2, \cdots, l$; $j = 1, 2, \cdots, k$, where l is an arbitrary positive integer.

[1] This is a slightly modified form of the definition in [1].

2. A known lemma. The main result will be established by a generalization of the following well known lemma:

LEMMA 1 (BERNOULLI). *Let $M = \{m\}$ and $M' = \{m'\}$ be two disjoint sets of measures. If there exists a set A in $\mathfrak{B}$ and a $\delta > 0$ such that*

$$| m(A) - m'(A) | > 2\delta$$

for all m in M and all m' in M', then, for any $\epsilon > 0$ given in advance, there exists an integer k and a set E in $\mathfrak{C}^k$ such that

$$m^k(E) < \epsilon \qquad \textit{for all } m \textit{ in } M$$

and

$$m'^k(E) > 1 - \epsilon \qquad \textit{for all } m' \textit{ in } M'.$$

This is an almost immediate consequence of Bernoulli's theorem, but for sake of completeness I include a proof.

PROOF. For any two integers n and r, $0 \leq r \leq n$, let $R(n, r)$ be the union of all regions in X^n defined by restricting r of the first n coordinates to A, and the remaining $n - r$ to $(X - A)$. For any fixed n, let for any measure μ

$$S_\mu^n = \bigcup_{\{r \mid |\frac{r}{n} - \mu(A)| \geq \delta\}} R(n, r).$$

(Here $\{t \mid T\}$ means the set of all t's which satisfy relation T, as customary.) Let $\epsilon > 0$ be given. By Bernoulli's theorem, there exists an integer $n(\epsilon)$ such that

$$\mu^n(S_\mu^n) = \sum_{\{r \mid |\frac{r}{n} - \mu(A)| \geq \delta\}} \binom{n}{r} [\mu(A)]^r [1 - \mu(A)]^{n-r} < \epsilon.$$

Let $E = \bigcap_{m \in M} S_m^n$. Since for any fixed n there are only a finite number of different S_m^n, E is in $\mathfrak{C}^n$ and

$$m^n(E) < \epsilon$$

for all m in M. Since for any m in M and m' in M',

$$\left| \frac{r}{n} - m(A) \right| \geq 2\delta - \left| m'(A) - \frac{r}{n} \right|,$$

we have $| (r/n) - m(A) | > \delta$ for all r satisfying $| m'(A) - (r/n) | < \delta$ for some m' in M'. Hence if x is in $(X^n - S_{m'}^n)$ for some m' in M', then x is in E and

$$m'^n(E) > m'^n(X^n - S_{m'}^n) > 1 - \epsilon$$

for all m' in M'.

In the special case $M = m$, $M' = m'$, this proves the statement in the introduction concerning simple hypotheses.

3. The main results.

DEFINITION. *Let $M = \{m\}$ and $M' = \{m'\}$ be two disjoint sets of measures. We shall say that they satisfy "Condition 1" if the following holds: M is the union*

of a finite number of its subsets M_i, $i = 1, 2, \cdots, k$, such that for every i there exist

(i) *a covering of M' by a finite number of its subsets M'_{ij}, $j = 1, 2, \cdots, h_i$,*

(ii) *a sequence of sets A_{ij} in $\mathfrak{B}$, $j = 1, 2, \cdots, h_i$, and*

(iii) *a $\delta > 0$ such that $| m(A_{ij}) - m'(A_{ij}) | > \delta$ for every m in M_i and every m' in M'_{ij}, $j = 1, 2, \cdots, h_i$; $i = 1, 2, \cdots, k$.*

Condition 1 is satisfied for instance if both M and M' contain only a finite number of measures.

LEMMA 2. *Let $M = \{m\}$ and $M' = \{m'\}$ be two disjoint sets of measures and assume that they satisfy Condition 1. Then for every $\epsilon > 0$, there exist an integer $n(\epsilon)$ and a set E in $\mathfrak{E}^{\cdot n}$ such that $m^n(E) < \epsilon$ for all m in M and $m'^n(E) > 1 - \epsilon$ for all m' in M'.*

PROOF. Assume first that $k = 1$. Then $M_i = M_1 = M$, and we can put $M'_{1j} = M'_j$, $h_1 = h$, $A_{ij} = A_j$. Condition 1 then states that $| m(A_j) - m'(A_j)| > \delta$ for every m in M and m' in M', $j = 1, 2, \cdots, h$. By Lemma 1, for any $\epsilon > 0$ there exists n_j and E_j in $\mathfrak{E}^{\cdot n_j}$ such that $m^{n_j}(E_j) < \epsilon/h$ for all m in M and $m'^{n_j}(E_j) > 1 - \epsilon/h$ for all m' in M'_j. Let $n = \max n_j$ and

$$E = \bigcup_{j=1}^{h} E_j \cdot X^{n-n_j}.$$

Then E is in $\mathfrak{E}^{\cdot n}$ and

$$m^n(E) \leq \sum_{j=1}^{h} m^n(E_j \cdot X^{n-n_j}) = \sum_{j=1}^{h} m^{n_j}(E_j) < \epsilon$$

for every m in M, and if m' is in any fixed M'_j,

$$m'^n(E) > m'^{n_j}(E_j) > 1 - \frac{\epsilon}{h} \geq 1 - \epsilon,$$

so that

$$m'^n(E) > 1 - \epsilon$$

for all m' in M'.

Now if $k > 1$, let us choose some $\bar{\epsilon}$, $0 < \bar{\epsilon} < \frac{1}{2}$, and apply the above argument to each M_i. We get $m^{n_i}(E_i) < \bar{\epsilon}$ for all m in M_i and $m'^{n_i}(E_i) > 1 - \bar{\epsilon}$ for all m' in M'. Hence

$$| m^{n_i}(E_i) - m'^{n_i}(E_i)| > 1 - 2\bar{\epsilon} > 0,$$

so that Condition 1 is satisfied with $k = 1$ and with the set $\{m'^{(\max n_i)}\}$ taking the place of M and the set $\{m^{(\max n_i)}\}$ taking the place of M'. It is easy to see that also in this case E still belongs to the field $\mathfrak{E}^{\cdot n}$.

If we do not require that the set E in the conclusion of Lemma 2 belong to $\mathfrak{E}^{\cdot n}$ but only that it belong to $\mathfrak{B}^{\cdot n}$, we can relax Condition 1 in the following obvious way:

THEOREM 1. *In order that two disjoint sets of measures $M = \{m\}$ and $M' = \{m'\}$ be such that for every $\epsilon > 0$ there be an integer n and a set B in $\mathfrak{B}^{\cdot n}$ for which*

$m^n(B) < \epsilon$ for every m in M and $m'^n(B) > 1 - \epsilon$ for every m' in M', it is necessary and sufficient that for some integer ν the sets $\{m^\nu\}$ and $\{m'^\nu\}$ satisfy Condition 1.

THEOREM 2. *Let $M = \{m_\theta\}$, $M' = \{m'_\tau\}$ be two disjoint sets of measures, $a \leq \theta \leq b$, $a' \leq \tau \leq b'$, where $[a, b]$ and $[a', b']$ are two disjoint, closed intervals of some finite-dimensional Euclidean space, and assume that, for each B in $\mathfrak{B}$, $m_\theta(B)$ and $m'_\tau(B)$ are continuous functions of θ and τ, respectively. Then for any $\epsilon > 0$ given in advance, there exist an integer $n(\epsilon)$ and a set E in $\mathfrak{E}^n$ such that $m_\theta^n(E) < \epsilon$ for all θ in $[a, b]$ and $m'^n_\tau(E) > 1 - \epsilon$ for all τ in $[a', b']$.*

PROOF. It is sufficient to prove that M and M' satisfy Condition 1. For any θ in $[a, b]$ and any τ in $[a', b']$, let $B_{\theta\tau}$ denote a set in $\mathfrak{B}$ for which

$$|m_\theta(B_{\theta\tau}) - m'_\tau(B_{\theta\tau})| > \epsilon_{\theta\tau} > 0. \tag{1}$$

(This is obviously possible.)

Let us now hold θ fixed. Because of the continuity of m'_τ, for every τ there exists a $\delta_{\theta\tau} > 0$ such that whenever $|\bar{\tau} - \tau| < \delta_{\theta\tau}$ then

$$|m'_{\bar{\tau}}(B_{\theta\tau}) - m'_\tau(B_{\theta\tau})| < \frac{\epsilon_{\theta\tau}}{3}. \tag{2}$$

Since $[a', b']$ is compact, it can be covered for each fixed θ by a finite subset of the open intervals $I_{\theta\tau} = (-\delta_{\theta\tau} + \tau, \tau + \delta_{\theta\tau})$, say $I_{\theta 1}, I_{\theta 2}, \cdots, I_{\theta h(\theta)}$, with midpoints $\tau_{\theta 1}, \tau_{\theta 2}, \cdots, \tau_{\theta h(\theta)}$. Denote the values of m'_τ, $h_{\theta\tau}$, $B_{\theta\tau}$, $\epsilon_{\theta\tau}$, $\delta_{\theta\tau}$ at $\tau = \tau_{\theta j}$ by $m'_{\theta j}$, $h_{\theta j}$, $B_{\theta j}$, $\epsilon_{\theta j}$, $\delta_{\theta j}$, respectively, for $j = 1, 2, \cdots, h(\theta)$.

Since m_θ is continuous in θ for all B, there exists a positive number ρ_θ such that whenever $|\bar{\theta} - \theta| < \rho_\theta$ then simultaneously for $j = 1, 2, \cdots, h(\theta)$

$$|m_\theta(B_{\theta j}) - m_{\bar{\theta}}(B_{\theta j})| < \tfrac{1}{3} \min_j \epsilon_{\theta j}, \tag{3}$$

and since $[a, b]$ is compact, it can be covered by a finite subset of the open intervals $L_\theta = (-\rho_\theta + \theta, \theta + \rho_\theta)$, say $L_i = (-\rho_i + \theta_i, \theta_i + \rho_i)$, $i = 1, 2, \cdots, k$. Let us denote the values of $\tau_{\theta j}$, $B_{\theta j}$, $h(\theta)$, $\epsilon_{\theta j}$, $\delta_{\theta j}$, $I_{\theta j}$ at $\theta = \theta_i$ by τ_{ij}, B_{ij}, h_i, ϵ_{ij}, δ_{ij}, I_{ij}, respectively. Then the sets $M_i = \{m_\theta \mid \theta \text{ in } L_i\}$, $i = 1, 2, \cdots, k$, cover M, and for each i the sets $|M'_{ij} = \{m'_\tau \mid \tau \text{ in } I_{ij}\}$, $j = 1, 2, \cdots, h_i$, cover M'. Furthermore it follows from (1), (2), and (3) that as long as θ is in M_i and τ in M'_{ij},

$$\begin{aligned} |m_\theta(B_{ij}) - m'_\tau(B_{ij})| &\geq |m_{\theta_i}(B_{ij}) - m'_{\tau_{ij}}(B_{ij})| - |m_{\theta_i}(B_{ij}) - m_\theta(B_{ij})| \\ &\quad - |m'_\tau(B_{ij}) - m'_{\tau_{ij}}(B_{ij})| > \epsilon_{ij} - \tfrac{1}{3} \min_j \epsilon_{ij} - \tfrac{1}{3} \epsilon_{ij} > \eta > 0, \\ &\qquad i = 1, 2, \cdots, k; \qquad j = 1, 2, \cdots, h_i, \end{aligned}$$

as we wanted to prove.

In the statistical terminology of the introduction, Theorem 2 may be restated as follows: Let H_0 be the hypothesis that the unknown distribution of some random variable belongs to a set of distributions M, and H_1 that it belongs to

another, M', where M and M' satisfy the assumptions of Theorem 2. Then there exists a uniformly consistent sequence of tests for testing H_0 against H_1.

4. An example. Let $F(t) = (1/\sqrt{2\pi})\, e^{-\frac{1}{2}(x-t)^2}$, $H_0 = F(0)$, $H_1 = \{F(t), 1 \le t \le 2\}$ Let

$$R_{n,i} = \left\{(x_1, x_2, \cdots, x_n) \,\middle|\, \frac{x_1 + x_2 + \cdots + x_n}{n} > c_i\right\},$$

$$i = 1, 2, \cdots; n = 1, 2, \cdots,$$

where c_i is determined so that $P(R_{n,i} \mid 0) = 1/i$, $P(S \mid t)$ denoting the probability of the region S when t is the true mean. Thus $R_{n,i}$ is the uniformly most powerful region of size $1/i$ in n-dimensional sample space for testing H_0 against H_1. The regions $R_{n,i}$ define a uniformly consistent test. A proof avoiding all computation is based on Theorem 2 as follows. Let $\epsilon > 0$ be given; find i such that $1/i < \epsilon$. By Theorem 2, there exists an N and a Borel set B in the N-dimensional sample space such that $P(B \mid 0) < 1/i$ and $P(B \mid t) > 1 - 1/i$ for $1 \le t \le 2$. Let W be a region in N-dimensional sample space covering B and such that $P(W \mid 0) = 1/i$. Then $P(W \mid 0) = P(R_{n,i} \mid 0) = 1/i < \epsilon$, and by the definition of $R_{N,i}$

$$P(R_{N,i} \mid t) \ge P(W \mid t) \ge P(B \mid t) > 1 - 1/i > 1 - \epsilon, \qquad 1 \le t \le 2.$$

It is a pleasure to express my best thanks to Professor J. Wolfowitz for calling my attention to the problem and to Professors J. von Neumann and H. Scheffé for their valuable suggestions.

REFERENCES

[1] A. Wald and J. Wolfowitz, "On a test whether two samples are from the same population," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 147–162.

[2] A. Wald, "Asymptotically most powerful tests of statistical hypotheses," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 1–19.

[3] A. Wald, "Some examples of asymptotically most powerful tests," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 396–408.

[4] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Trans. Am. Math. Soc.*, Vol. 54 (1943), pp. 426–482.

[5] R. von Mises, "On the problem of testing hypotheses," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 238–252.

NONPARAMETRIC ESTIMATION IV

By D. A. S. Fraser and R. Wormleighton[1]

University of Toronto

1. Summary. In the three papers, [1], [2], [3], entitled "Nonparametric estimation", Scheffé and Tukey generalized previous results on tolerance regions and extended them to cover all continuous and discontinuous distribution functions. This note contains four comments arising from these papers: first, on a method for giving bounds to the confidence level in the discontinuous case which can lower the probability that the end points need to have part, a random variable, of their probability neglected to maintain the given confidence level; second, on a correction of a statement of results in [2]; third, on a proof in [2] requiring a further statement; fourth, a necessary restatement of theorems in [3].

2. Bounds for the confidence level. In paper [1] Scheffé and Tukey extend the theory of tolerance regions to the one dimensional discontinuous case, and obtain the following statement:

$$Pr\{C_{(p)-0,(q)+0} \geqq \beta\} \geqq 1 - \alpha_{q-p} \geqq Pr\{C_{(p)+0,(q)-0} \geqq \beta\}, \tag{2.1}$$

where $C_{(p)\pm 0,(q)\mp 0}$ are respectively the coverages of the open and closed intervals with end points the pth and qth order statistics z_p and $z_q(q > p)$ of a sample of n from a distribution and α_{q-p} is the incomplete Beta function $I_\beta(q - p, n - p + q + 1)$.

This statement implies the following statements:

$$1 - \alpha_{q-p+2} \geqq Pr\{C_{(p-1)+0,(q+1)-0} \geqq \beta\},$$

$$Pr\{C_{(p+1)-0,(q-1)+0} \geqq \beta\} \geqq 1 - \alpha_{q-p-2}\,.$$

This suggests giving bounds for the confidence levels of the tolerance regions of statement (2.1).

Let us consider the one dimensional representation theorem with its "inverse probability integral transformation". This transformation labelled $g_F(x^*)$ mapped x^* with a uniform distribution into x with the given distribution represented by $F(x)$. z^* and z refer to the corresponding order statistics. Take any interval on the range of the uniform distribution whose end points lie respectively in the closed intervals (z^*_{p-1}, z^*_p) and (z^*_q, z^*_{q+1}). The confidence level, that the coverage of this interval is at least β, lies between $1 - \alpha_{q-p}$ and $1 - \alpha_{q-p+2}$. Apply the mapping $g_F(x^*)$. The confidence level lies between $1 - \alpha_{q-p}$ and $1 - \alpha_{q-p+2}$ that the following coverage is greater than or equal to β: $C_{(p)-0,(q)+0}$, if z_p is distinct from z_{p-1} and z_q is distinct from z_{q+1}; $C_{(p)-0,(q)-0}$ + "fraction" of the coverage of z_q if z_p is distinct from z_{p-1} and z_q identical to z_{q+1}; and similarly for the other two possible cases. The "fraction" (a number between 0 and 1) can

[1] The authors wish to thank Professor John Tukey for suggesting Definitions 5.2 and 5.3.

be considered as a random variable as determined by the above mapping or as a fixed value since the relation must be true for at least one fixed value for any given distribution and integers p and q. In either case it is unknown to the practitioner and the interpretation would be unimportant.

Similarly we obtain the following result: The confidence level lies between $1 - \alpha_{q-p}$ and $1 - \alpha_{q-p+r+s}$ that the following coverage is greater than or equal to β: $C_{(p)-0,(q)+0}$, if z_p is distinct from z_{p-r} and z_q is distinct from z_{q+s}; $C_{(p)-0,(q)-0}$ + "fraction" of the coverage of z_q, if z_p is distinct from z_{p-r} and z_q is identical to z_{q+s}; and similarly for the other two cases.

The open interval can be treated in a similar manner.

The application of these results would be for the practitioner who was familiar with the type of data he was to receive and realized that perhaps two or three order statistics would be tied on one or perhaps both tails. He would then choose r and s to give as tight control of the confidence level consistent with a reasonable determinacy in the tolerance interval (the probability being small that the coverage should be considered as including only part instead of all of the coverage of the end points).

These results also generalize to the multivariate case with little alteration. For example consider the following result which would correspond to the closed interval case above. The confidence level lies between $1 - \alpha_{q-p}$ and $1 - \alpha_{q-p+r}$ that the following coverage is greater than or equal to β: cov $\{\bar{B}_\lambda\}$, if $\bar{B}_\lambda$ is contained in $B_{\lambda+\mu}$ where μ consists of r of the integers $(1, 2, \cdots, n + 1)$ which are not contained in λ; cov $\{B_\lambda\}$ + "fraction" of (cov $\{\bar{B}_\lambda\}$ − cov $\{B_\lambda\}$), if $\bar{B}_\lambda$ is not contained in $B_{\lambda+\mu}$. Here the "fraction" can be considered a random variable or fixed, in either case unknown to the practitioner (λ containing $q - p$ integers).

In formulating the above generalization, attention was drawn to the fact that the block groups did not form a proper sequence as λ was increased. By the following counter example the theorems in [3] are seen to be incorrect using the given definition of block groups. Rectifying definitions are presented in Section 5.

Following the notation of [3], let

$$\begin{aligned} \varphi_1(x, y) &= y, \\ \varphi_2(x, y) &= x, \\ \varphi_3(x, y) &= -y, \\ \varphi_4(x, y) &= -x, \end{aligned}$$

and $p_i = i (i = 1, 2, 3, 4)$.

Consider the distribution $F(x, y) = \epsilon(x)\epsilon(y)$ where $\epsilon(x)$ is defined by

$$\begin{aligned} \epsilon(x) &= 0, \quad x < 0, \\ &= 1, \quad x \geqq 0. \end{aligned}$$

Take a sample of $n \geqq 6$ from this distribution; the sample values will all be (0, 0) with probability one.

$$
\begin{aligned}
S_1 &= \{(x, y) \mid y > 0 \text{ or } y = 0,\ x > 0\},\\
T_1 &= \{(0, 0)\},\\
S_2 &= \{(x, y) \mid y < 0,\ x \geqq 0\},\\
T_2 &= \text{Null set},\\
S_3 &= \{(x, y) \mid x < 0,\ y \leqq 0\},\\
T_3 &= \text{Null set},\\
S_4 &= \text{Null set}.
\end{aligned}
$$

The corresponding coverages are respectively 0, 1, 0, 0, 0, 0, 0.

Taking $\lambda = \{3\}$ we find by the definition of block groups that

$$\bar{B}_\lambda = T_2 \cup S_3 \cup T_3\,, \qquad \bar{C}_\lambda = 0$$

with probability 1. Thus $Pr\{\bar{C}_\lambda < a\} = 1 \lneqq I_a(1, n)$.

Taking $\lambda = \{1, 2\}$ we have $B_\lambda = S_1 \cup T_1 \cup S_2$, and $C_\lambda = 1$ with probability 1. Thus $Pr\{C_\lambda < a\} = 0 \gneqq I_a(2, n - 1)$.

The proof in [3] is in error on page 39, the seventh line from the bottom.

3. Correction of a statement of results. In paper [2] on page 536 a "Statement of Results for Measure Theorists" is given. The theorem B_{n+1} should read: Hold the n functions $\varphi_1, \varphi_2, \cdots, \varphi_n$ and the probability measure fixed, then T^n is mapped on B_n and the power measure μ^n is carried by that mapping into a measure of B_n. This measure is always $n!/\sqrt{n+1}$ times Lebesgue measure.

4. A proof; a further statement. In the proof on page 537 of paper [3], the problem is to show that the distribution of $n - m$ variates is the same when obtained by two methods of calculation; more particularly, to show that, given that in a sample of n, one value falls in each of the sets $A_1, A_2, \cdots, A_m$ and the remaining $n - m$ fall in B, then the distribution of the $n - m$ in B is that of a sample of $n - m$ restricted to B. The statement is made that the probability of the above, and in addition that the $n - m$ falling in B, fall in $R \subset B$, is

$$\frac{n!}{(n-m)!}\ \mu(A_1)\mu(A_2) \cdots \mu(A_m)\mu^{n-m}(B)$$

times the probability that a sample of $n - m$ restricted to B falls in R. To show that the distributions are identical, a further statement is needed: that for one variate in each A_i, for p falling in R, and $n - m - p$ in $B - R$, then the probability is equal to

$$\frac{n!}{(n-m)!}\ \mu(A_1) \cdots \mu(A_m)\mu^{n-m}(B)$$

times the probability that a sample of $n - m$ restricted to B divides p into R and $n - m - p$ into $B - R$.

5. Restatement of theorems in [3]. As has been noted in Section 2 above, Theorems $A^*_{m|n+1}$ and B^*_{n+1} fail when actual ties (coincident points) occur. The following redefinition of the block groups overcomes this difficulty and the proof follows as given in [3].

Define S_i and T_i as in (4.1) in [3].

Definition 5.1. *Let $\bar{S}_i$ be given by the definition for S_i where $<$ is replaced by $\leqq$ and $>$ is replaced by $\geqq$.*

Definition 5.2. *The block group B_λ consists of the union of all S_i with i in λ and all T_j not contained in any $\bar{S}_i$ with i not in λ.*

Definition 5.3. *The closed block group $\bar{B}_\lambda$ consists of the union of all S_i with i in λ and all T_j contained in any $\bar{S}_i$ with i in λ.*

Using the above definitions, Theorems $A^*_{m|n+1}$ and B^*_{n+1} follow provided the "m-system of functions" is chosen so that all T_i are reduced to points.

A more general definition of block groups which will cover cases where the "m-system of functions" does not reduce all cuts to points and which is identical to that of (5.2) and (5.3) when all cuts are necessarily points is given by (5.4) and (5.5).

Definition 5.4. *The closed block group $\bar{B}_\lambda$ consists of the union of all $\bar{S}_i$ with i in λ.*

Definition 5.5. *The block group B_λ consists of the complement of $\bar{B}_{C(\lambda)}$ where $C(\lambda)$ is the complement of λ with respect to the integers $(1, 2, \cdots, n+1)$.*

According to the representation theorem in [3], we have a continuous joint distribution of variates $U_1, U_2, \cdots, U_m$. By means of monotone functions $g_1(U_1), \cdots, g_m(U_m)$ this continuous distribution is mapped into a discontinuous distribution identical to the distribution of $\psi_1(w_1), \cdots, \psi_m(w_m)$.

$$\begin{aligned}
\text{Let } S_1' &= \{(U_1, \cdots, U_m) \mid U_1 > u_1(i_{(1)})\},\\
S_2' &= \{(U_1, \cdots, U_m) \mid U_1 < u_1(i_{(1)}),\ U_2 > u_2(i_{(2)})\},\\
&\ \ \vdots\\
S_m' &= \{(U_1, \cdots, U_m) \mid U_1 < u_1(i_{(1)}), \cdots, U_m > u_m(i_{(m)})\},\\
S_{m|n+1}' &= \{(U_1, \cdots, U_m) \mid U_1 < u_1(i_{(1)}), \cdots, U_m < u_m(i_{(m)})\}.
\end{aligned}$$

Also we have:

$$\begin{aligned}
S_1^* &= \{g_1(U_1), \cdots, g_m(U_m) \mid g_1(U_1) > g_1(u_1(i_{(1)}))\},\\
S_2^* &= \{g_1(U_1), \cdots, g_m(U_m) \mid g_1(U_1) < g_1(u_1(i_{(1)})),\ g_2(U_2) > g_2(u_2(i_{(2)}))\},\\
&\ \ \vdots\\
S_{m|n+1}^* &= \{g_1(U_1), \cdots, g_m(U_m) \mid g_1(U_1) < g_1(u_1(i_{(1)})), \cdots, g_m(U_m) < g_m(u_m(i_{(m)}))\},
\end{aligned}$$

and $\bar{S}_1^*$, $\bar{S}_2^*$, etc., are defined as S_1^*, S_2^*, etc., where $<$ is replaced by $\leqq$ and $>$ is replaced by $\geqq$.

Consider now the inverse mapping of the sets $S_1^*, S_2^*, \cdots$ and $\bar{S}_1^*, \bar{S}_2^*, \cdots$ into the space of $(U_1, U_2, \cdots, U_m)$. We shall have

$$g^{-1}(S_i^*) \subset S_i' \subset g^{-1}(\bar{S}_i^*)$$

because

$$g_i(U_i) > g_i(a) \rightarrow U_i > a \rightarrow g_i(U_i) \geqq g_i(a).$$

Thus we have the following inequality for the corresponding coverages:

$$\text{cov}\ (S_i^*) \leqq \text{cov}\ (S_i') \leqq \text{cov}\ (\bar{S}_i^*).$$

The proof follows directly from this relation as in section (9) of [3].

REFERENCES

[1] H. Scheffé and J. W. Tukey, "Nonparametric estimation I," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 187–192.
[2] J. W. Tukey, "Nonparametric estimation II," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 529–539.
[3] J. W. Tukey, "Nonparametric estimation III," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 30–39.

NOTES

CONDITIONAL EXPECTATION AND THE EFFICIENCY OF ESTIMATES

BY PAUL G. HOEL

University of California, Los Angeles

1. Summary. A probability density function, $f(x; \theta)$, is considered for which there exists a sufficient statistic. It is shown, under certain regularity conditions on the family of distributions and on the class of estimates, that if there exists an unbiased sufficient estimate of θ, it will be unique. This result is used to show that when the regularity conditions are satisfied, the method of Blackwell for improving an unbiased estimate of θ merely yields a natural estimate.

2. Distribution of a sufficient statistic. For the purpose of proving the uniqueness of an unbiased sufficient estimate of θ, it is helpful to find the form of the probability density function of a particular sufficient statistic from a knowledge of the form of $f(x; \theta)$. It has been shown [1], [2] under different sets of assumptions that when a sufficient statistic exists, $f(x; \theta)$ must possess the functional form

$$f(x; \theta) = \exp\,[g(\theta) + h(\theta)r(x) + s(x)], \tag{1}$$

provided the range of x does not depend on θ.

ASSUMPTION 1. *It will be assumed that $f(x; \theta)$ has the form given by* (1).

Koopman [1] proves (1) under the assumption that $f(x; \theta)$ is analytic. Pitman [2] assumes only that $\partial^2 f(x; \theta)/\partial x \partial\theta$ exists but adds a differentiability condition on the density function of the sufficient statistic that is assumed to exist.

Now consider the distribution of a particular statistic. From (1), the probability density function for a random sample is

$$f(x_1, \cdots, x_n; \theta) = \exp\,[ng(\theta) + h(\theta)\sum_{i=1}^{n} r(x_i) + \sum_{i=1}^{n} s(x_i)]. \tag{2}$$

The particular statistic to be considered here is $z = \Sigma r(x_i)$. From a lemma of Lehmann and Scheffé [3], mild regularity conditions will insure the existence of a transformation to new variables $z, t_2, \cdots, t_n$ such that the density function of the new variables may be expressed in the form

$$F(z, t_2, \cdots, t_n; \theta) = f(x_1, \cdots, x_n; \theta)/|\,J\,|, \tag{3}$$

where J is the Jacobian of the transformation and where the x's are replaced by their expressions in terms of the new variables. The essential regularity conditions here are that $r'(x) \neq 0$, except possibly on a set of measure zero, and that $r'(x)$ is continuous.

ASSUMPTION 2. *It will be assumed that the lemma conditions are satisfied.*

If the assumptions made in [1] had been employed, these restrictions on $r(x)$ would have been satisfied because then $r(x)$ would be analytic.

If (2) is substituted in (3), and then (3) is integrated over the range of the variables $t_2, \cdots, t_n$, the density function of z will be obtained in the form

$$p(z; \theta) = L(z) \exp [ng(\theta) + h(\theta)z], \tag{4}$$

because $\exp [\Sigma s(x_i)]/| J |$ does not involve θ and thus its integral over the range of the t's will be a function of z only.

It is easily seen that z is a sufficient statistic because it suffices to show that $f(x_1, \cdots, x_n; \theta)/p(z; \theta)$ is independent of θ. From (2) and (4) it is clear that this ratio is independent of θ.

3. Relationship of sufficient estimates. If an estimate of θ is understood to be a single-valued function of the random variables $x_1, \cdots, x_n$, and if certain derivatives exist, then any sufficient estimate will be a function of z. For, let y be any sufficient estimate. Then

$$f(x_1, \cdots, x_n; \theta) = G(y; \theta)H(x_1, \cdots, x_n), \tag{5}$$

where $G(y; \theta)$ is the density function of y. If $\partial \log f/\partial\theta$ is calculated for both (2) and (5) and if the results are equated, it will follow that

$$\frac{\partial \log G\,(y; \theta)}{\partial\theta} = ng'(\theta) + h'(\theta)z.$$

This result shows that z is a single-valued function of y, when the derivatives exist. Conversely, since only single-valued functions of the variables $x_1, \cdots, x_n$ are considered as estimates, y will be a single-valued function of z. If the relationship is $z = T(y)$ and $y = T^{-1}(z)$ is multiple-valued, y will be defined only on one branch.

ASSUMPTION 3. *It will be assumed that $g'(\theta)$, $h'(\theta)$, and $\partial G(y; \theta)/\partial\theta$ exist for some value of θ.*

The restriction that $g'(\theta)$ and $h'(\theta)$ should exist would be satisfied if the assumptions made in either [1] or [2] to arrive at formula (1) had been made instead of Assumption 1. The restriction that $\partial G(y; \theta)/\partial\theta$ should exist is a weak restriction on the class of sufficient estimates, y, being considered.

4. Uniqueness. Suppose there exists an unbiased sufficient estimate of θ. From the preceding section it will be a function of z, say $w(z)$. Then, from (4),

$$\int_{-\infty}^{\infty} w(z)L(z)e^{ng(\theta)+h(\theta)z}\,dz = \theta. \tag{6}$$

Now suppose that there were two such estimates, say $w_1(z)$ and $w_2(z)$. Then, letting $\alpha = h(\theta)$, (6) would yield the following relationship:

$$\int_{-\infty}^{\infty} w_1(z)L(z)e^{\alpha z}dz = \int_{-\infty}^{\infty} w_2(z)L(z)e^{\alpha z}dz. \tag{7}$$

But from the theory of Laplace transforms ([4], p. 244), it follows that (7) implies that $w_1(z) = w_2(z)$ except on sets of measure zero, provided that $w_1(z)L(z)$ and $w_2(z)L(z)$ are integrable in every finite interval and provided that (7) holds for some interval of values of α. It is easily seen that the existence of (6) for all admissible values of θ insures the integrability condition. If $f(x; \theta)$ is defined for an interval of values of θ, (6) will hold for such an interval. Since, from Assumption 3, $\alpha = h(\theta)$ is continuous and from (1) is obviously not constant, α must exist for an interval of values also, and hence (7) will hold for such an interval.

ASSUMPTION 4. *It will be assumed that $f(x; \theta)$ is defined for some interval of values of θ.*

The discussions of the preceding sections may be summarized in the following theorem.

THEOREM. *If Assumptions 1, 2, 3, and 4 are satisfied and if there exists an unbiased sufficient estimate of θ, it will be the unique such estimate.*

5. Efficiency. Let t be any unbiased estimate of θ and let u be any sufficient estimate of θ. Then Blackwell [5] has shown that

$$v = E[t \mid u], \tag{8}$$

which is the conditional expected value of t for u fixed, determines an unbiased estimate of θ whose variance cannot exceed that of t. If t is not a function of u, the variance of v will be less than that of t.

This device for improving the efficiency of an unbiased estimate appears promising. However, since v is a function of u and thus is a sufficient unbiased estimate of θ, and since it has been shown that, subject to mild regularity conditions, any unbiased sufficient estimate is unique, this device will merely yield the unique unbiased sufficient estimate. Since the statistic z may be found by inspecting $f(x; \theta)$, it suffices to find the function $w(z)$ which is unbiased in order to obtain the desired estimate, when such an estimate exists. From (8), the existence of an unbiased estimate insures the existence of an unbiased sufficient estimate. Since, from (2), the maximum likelihood estimate of θ is a function of z, a natural method for finding this unique estimate when it exists is to first find the maximum likelihood estimate and then, if necessary, determine what function of it will be unbiased. Formula (8) with u chosen as z could also be used to find this unique estimate.

REFERENCES

[1] B. O. KOOPMAN, "On distributions admitting a sufficient statistic," *Trans. Am. Math. Soc.*, Vol. 39 (1936), p. 399.
[2] E. J. G. PITMAN, "Sufficient statistics and intrinsic accuracy," *Proc. Cambridge Philos. Soc.*, Vol. 32 (1936), p. 567.
[3] E. L. LEHMANN AND H. SCHEFFÉ, "On the problem of similar regions," *Proc. Nat. Acad. Sci.*, Vol. 33 (1947), p. 383.
[4] D. V. WIDDER, *The Laplace Transform*, Princeton University Press, 1946.
[5] D. BLACKWELL, "Conditional expectation and unbiased sequential estimation," *Annals of Math. Stat.*, Vol. 18 (1947), p. 105.

LINEAR TRANSFORMATIONS AND THE PRODUCT-MOMENT MATRIX

By K. Nagabhushanam

Institute of Mathematical Statistics, Stockholm

Using linear transformations G. Rasch has deduced Wishart's distribution in his paper on "A Functional Equation for Wishart's Distribution" (*Annals of Math. Stat.*, Vol. 19 (1948), pp. 262–266). This note is of the nature of some observations on the Jacobian of the transformation induced by a linear transformation of coordinates with constant coefficients in the distinct elements of the product-moment matrix of a sample of n vectors, each of k components, drawn from a universe of a normal k-variate distribution with zero means. If the r-th vector of the drawn sample has components $(x_i^{(r)})$, $r = 1, 2, \cdots, n$, $i = 1, 2, \cdots, k$, the sum of the products of the i-th and j-th components of each of the n vectors is denoted by

$$M_{ij} = \sum_{r=1}^{n} x_i^{(r)} x_j^{(r)}.$$

Let the variables of the vector variate be $x_1, x_2, \cdots, x_k$, or shortly (x), and let a nonsingular (i.e., reversible) linear transformation of the variables with constant coefficients be made from (x) to (y), viz.,

$$x_r = \sum_{i=1}^{k} a_{ri} y_i \qquad (r = 1, 2, \cdots, k).$$

The distinct elements $M_{11}, M_{12}, \cdots, M_{1k}, M_{22}, M_{23}, \cdots, M_{kk}$ undergo a consequential or induced transformation which is also linear in terms of the corresponding elements of the product-moment matrix $\| M'_{ij} \|$ of the same n-vector sample in the coordinates (y).

Let the matrix of the coefficients of the induced transformation which is also the matrix of partial derivatives in this case be denoted by $\| J \|$, and let its determinant which is the Jacobian of the transformation be denoted by J. The elements of $\| J \|$ are functions of the elements of $\| a_{ri} \|$. When $\| a_{ri} \|$ is in the diagonal form, so is also $\| J \|$ with elements $a_{11}a_{11}, a_{11}a_{22}, \cdots, a_{11}a_{kk}, a_{22}a_{22}, a_{22}a_{33}, \cdots, a_{kk}a_{kk}$. In this case we have $J = A^{k+1}$, where A stands for the determinant of $\| a_{ri} \|$ which is nonzero on account of the nonsingularity of the transformation considered. It is then natural to ask the question whether it can be asserted that the same relationship holds even when the matrix $\| a_{ri} \|$ is not in the diagonal form or reducible to it. The answer is in the affirmative, the result being a particular case of the following theorem of Escherich (see C. C. Macduffee, *Theory of Matrices*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Julius Springer, Berlin, 1933, p. 86, theorem 44.3): *The determinant of the m-th power-matrix*[1] *with a nonvanishing determinant A is $A^{(m+k-1)C_k}$, where*

[1] It may be noted that the m-th power-matrix mentioned in Escherich's theorem is not the matrix multiplied by itself by the ordinary matrix multiplication. For its definition see Macduffee, loc. cit., p. 85.

k is the order of the square matrix whose determinant is A. Here $\| J \|$ is the second power-matrix, and so we have in all cases $J = A^{k+1}$.

Due to the importance of this in connection with Wishart's distribution, the following observations are of special interest.

(1) When the latent roots of the matrix $\| a_{ri} \|$ are all distinct, it is reducible to the diagonal form, and it has been shown by Rasch that in such a case $J = A^{k+1}$, and the method of analytic continuation has been suggested as a means of establishing the same result when not all the latent roots are distinct. We shall show this here by a consideration of the limit of a polynomial function.

Let us suppose that some of the latent roots are repeated. Let now a suitable number of the elements of the matrix be replaced by neighbouring values, i.e., by $a_{ri} + \epsilon\eta_{ri}$, so that the latent roots of the altered matrix, $\| a'_{ri} \|$ say, are all distinct. (This is always possible as there are an adequate or more than adequate number of elements for this purpose.) We shall now consider a linear transformation from the variables (x) to the variables (y) with $\| a'_{ri} \|$ as the transformation matrix. Using primes to denote the corresponding quantities in relation to the new transformation of the variables, we have $J' = (A')^{k+1}$, since $\| a'_{ri} \|$ is reducible to the diagonal form. Further J', expressed as an expansion in its elements, is a polynomial function whose continuity properties yield by proceeding to the limit

$$\begin{aligned} J = \lim_{\epsilon \to 0} J' &= \lim_{\epsilon \to 0} (A')^{k+1} \\ &= (\lim_{\epsilon \to 0} A')^{k+1} \\ &= A^{k+1}. \end{aligned}$$

(2) The following situation is sometimes met with. Consequent on the linearity of the transformation of the variables with constant coefficients we have

$$dx_r = \sum_{i=1}^{k} a_{ri}\, dy_i \qquad (r = 1, 2, \cdots k)$$

which gives

$$d(x_i^{(r)} x_j^{(r)}) = \sum_{l=1}^{k} \sum_{m=1}^{k} a_{il}\, a_{jm}\, d(y_l^{(r)} y_m^{(r)}).$$

Summing over the n values of r, we have dM_{ij} transformed in terms of (dM'_{lm}) with the same coefficients as those that appear in the transformation of $dx_i\, dx_j$ in terms of $(dy_l\, dy_m)$. Also we know that

$$\begin{aligned} &dM_{11}\, dM_{12} \cdots dM_{1k}\, dM_{22}\, dM_{23} \cdots dM_{kk} \\ &\qquad = J \cdot dM'_{11}\, dM'_{12} \cdots dM'_{1k}\, dM'_{22}\, dM'_{23} \cdots dM'_{kk}. \end{aligned}$$

By the sameness of coefficients of transformation we can write

$$\begin{aligned} &dx_1\, dx_1\, dx_1\, dx_2 \cdots dx_1\, dx_k\, dx_2\, dx_2\, dx_2\, dx_3 \cdots dx_k\, dx_k \\ &\qquad = J \cdot dy_1\, dy_1\, dy_1\, dy_2 \cdots dy_1\, dy_k\, dy_2\, dy_2\, dy_2\, dy_3 \cdots dy_k\, dy_k\,; \end{aligned}$$

that is,

$$(dx_1 dx_2 \cdots dx_k)^{k+1} = J \cdot (dy_1 dy_2 \cdots dy_k)^{k+1}.$$

But $dx_1 dx_2 \cdots dx_k = A \cdot dy_1 dy_2 \cdots dy_k$, so that $J = A^{k+1}$. These formal operations show that in this case the differentials when multiplied in the usual way work like the determinants they signify.

(3) After obtaining that $J = A^{k+1}$ Rasch's functional equation is seen to hold good.

(4) When the constant in Wishart's distribution is evaluated in Rasch's notation using H. Cramér's method (see *Mathematical Methods of Statistics*, Princeton University Press, 1946, pp. 390–393), it will be found that a power of n is missing in the numerator. This is due to the fact that we have not considered the estimate $1/n \, \| M_{ij} \|$ but worked with $\| M_{ij} \|$.

I thank Professor H. Cramér and Mr. G. Blom for their interest in this work.

A NOTE ON A TWO SAMPLE TEST[1]

By Frank J. Massey, Jr.

University of Oregon

1. Summary. Mood ([1], p. 394) discusses a test for the hypothesis that two samples come from populations having the same continuous cumulative frequency distributions. It consists of arranging the observations from the two samples in a single group in order of size and then comparing the numbers in the two samples above the median. This technique is extended to using several order statistics from the combined samples, and to the case of several samples. The test is nonparametric and might be a good substitute for the single variable of classification analysis of variance in cases of doubtful normality. The application of the test would be the same as in Mood ([1], p. 398) except that there would be more than two rows in the table.

2. The distribution function.

Suppose we have p populations all having the same continuous cumulative distribution function $F(x)$. Let X_{ij} ($i = 1, 2 \cdots, p; j = 1, 2 \cdots, n_i$) be the jth observation in a sample of size n_i from the ith population. Let $\sum_{i=1}^{p} n_i = N$.

Arrange these N observations in a single series according to size and rename them $z_1 \leq z_2 \leq \cdots \leq z_N$. We choose $k - 1$ of the z values, for example, $z_{\alpha_1}, z_{\alpha_2}, \cdots, z_{\alpha_{k-1}}$ (the α_i are integers and $1 \leq \alpha_1 < \alpha_2 < \cdots \leq N$). Denote by m_{ij} the number of observations X_{ih} such that $z_{\alpha_{j-1}} < X_{ih} \leq z_{\alpha_j}$ for $j = 2, 3, \cdots, k - 1$, by m_{i0} the number of $X_{ih} \leq z_{\alpha_1}$, and by m_{ik} the number of $X_{ih} > z_{\alpha_{k-1}}$. These can be illustrated by the following table.

[1] This paper sponsored in part by the Office of Naval Research, Contract N6-onr-218/IV, Project NR 042 063.

1st sample	2nd sample	$\cdots$	pth sample	
m_{1k}	m_{2k}	$\cdots$	m_{pk}	$\sum_{i=1}^{p} m_{ik} = N - \alpha_{k-1}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
m_{13}	m_{23}	$\cdots$	m_{p3}	$\sum_{i=1}^{p} m_{i3} = \alpha_3 - \alpha_2$
m_{12}	m_{22}	$\cdots$	m_{p2}	$\sum_{i=1}^{p} m_{i2} = \alpha_2 - \alpha_1$
m_{11}	m_{21}	$\cdots$	m_{p1}	$\sum_{i=1}^{p} m_{i1} = \alpha_1$
$\sum_{j=1}^{k} m_{1j} = n_1$	$\sum_{j=1}^{k} m_{2j} = n_2$	$\cdots$	$\sum_{j=1}^{k} m_{pj} = n_p$	

The joint distribution of the m_{ij} and z_i can be written as

$$\frac{\prod_{i=1}^{p} n_i!}{\prod_{i=1}^{p}\prod_{j=1}^{k} m_{ij}!} F(z_1)^{\alpha_1-1}\,dF(z_1)[F(z_2) - F(z_1)]^{\alpha_2-\alpha_1-1}\,dF(z_2)\cdots$$
$$\cdot[F(z_{k-1}) - F(z_{k-2})]^{\alpha_{k-1}-\alpha_{k-2}-1}[1 - F(z_{k-1})]^{N-\alpha_k}\,dF(z_{k-1})$$
$$\cdot\Sigma m_{\beta_1 1} m_{\beta_2 2} m_{\beta_3 3}\cdots m_{\beta_{k-1} k-1},$$

where the sum runs over all possible sets of values of $\beta_i = 1, 2, 3, \cdots p$. It is easy to show that this sum is equal to the product

$$\alpha_1(\alpha_2 - \alpha_1)(\alpha_3 - \alpha_2)\cdots(\alpha_{k-1} - \alpha_{k-2}).$$

Now the joint distribution of the m_{ij} is obtained by integrating the z's over their entire range $z_i \leq z_{i+1}$. This is a type of Dirichlet integral and we get

$$f(m_{11}\cdots m_{pk}) = \frac{\prod_{i=1}^{p} n_i!\prod_{j=1}^{k}(\alpha_j - \alpha_{j-1})!}{N!\prod_{i=1}^{p}\prod_{j=1}^{k} m_{ij}!},$$

where $\alpha_k = N$, $\alpha_0 = 0$. This is the distribution of cell frequencies in a p by k contingency table with all totals fixed when there is independence (see [1], p. 278) and thus, for large values of m_{ij} at least, the usual chi-square test can be used.

REFERENCE

[1] A. M. Mood, *Introduction to the Theory of Statistics*, McGraw-Hill Book Co., 1950.

AN OMISSION IN NORTON'S LIST OF 7 × 7 SQUARES

By Albert Sade

Marseille, France

1. In a previous paper the value 16,942,080 for the number of reduced 7 × 7 squares was obtained by the author by an exhaustive method, subject to a strict control ([4], Section 20). This number exceeds Norton's ([2], Table on p. 290) by 14,112. An attempt was made in Section 21 of [4] to show that this discrepancy in the total number does not affect Norton's conjecture ([2], p. 291) that the 146 species represent the whole of the universe of 7 × 7 Latin squares. However, R. A. Fisher has informed the author that the discrepancy cannot be explained away in this manner. It has therefore to be attributed to a gap in Norton's list.

2. Now, a 147th species containing 14,112 squares can arise only from an automorph type through an operator of the order 5^k. It is easy to construct a matrix Q corresponding to such an operator as, for example, $T = (34567)^3$. Here the cycle (34567) signifies a permutation [1] of columns, a permutation of rows and a substitution of elements.

The first two rows of Q are respectively identical with the first two columns and define the substitution (12) (34567). In the remaining 5 × 5 squares, it is necessary that the elements of the broken diagonals follow in the natural cyclic order, except the numbers 1 and 2, which each form a broken diagonal.

The square is given below:

$$Q = \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 1 & 4 & 5 & 6 & 7 & 3 \\ 3 & 4 & 5 & 7 & 1 & 2 & 6 \\ 4 & 5 & 7 & 6 & 3 & 1 & 2 \\ 5 & 6 & 2 & 3 & 7 & 4 & 1 \\ 6 & 7 & 1 & 2 & 4 & 3 & 5 \\ 7 & 3 & 6 & 1 & 2 & 5 & 4. \end{matrix}$$

3. On replacing each row of Q by the conjugate permutation and rotating the figure through an angle of 180° about the diagonal, we obtain the square

$$
R = \begin{matrix}
1 & 2 & 5 & 6 & 7 & 3 & 4 \\
2 & 1 & 6 & 7 & 3 & 4 & 5 \\
3 & 7 & 1 & 5 & 4 & 6 & 2 \\
4 & 3 & 2 & 1 & 6 & 5 & 7 \\
5 & 4 & 3 & 2 & 1 & 7 & 6 \\
6 & 5 & 7 & 4 & 2 & 1 & 3 \\
7 & 6 & 4 & 3 & 5 & 2 & 1.
\end{matrix}
$$

It is easy to verify the equality

$$R(12\cdot 36475)(12\cdot 34567) = Q,$$

in which the first factor is a permutation of columns and the second a substitution of numbers.

Thus the number of reduced squares produced by Q is

$$6(7!)^2/(3\cdot 5\cdot 7!\,6!) = 14{,}112,$$

which is precisely the difference mentioned in Section 1.

4. The reversal of the unique intercalate 12 in Q gives a square S isomorphic with Q, and on interchanging rows and columns and the numbers 1 and 2 in S, we obtain Q again. Thus S is one of the 14,112 squares considered in Section 3 and does not give a new species. Therefore, Norton's conjecture ([2], p. 291) "that they can be enumerated by an exhaustive reversal of intercalates" is not borne out, at least for species with one intercalate. This assumption was founded on its truth for 6×6 squares; but it is to be expected that the classification of $n \times n$ squares would become more complicated with increasing n.

5. On the contrary, the conclusion of S. G. Ghurye [3] is confirmed, for the square Q possesses a different "$I - A$" from those of other species.

REFERENCES

[1] T. B. Sprague, "A new algebra, by means of which permutations can be transformed in a variety of ways, and their properties investigated," *Trans. Roy. Soc. Edinburgh*, Vol. 37 (1892), pp. 399–411.

[2] H. W. Norton, "The 7 × 7 squares," *Annals of Eugenics*, Vol. 9 (1939), pp. 269–307.

[3] S. G. Ghurye, "A characteristic of species of 7 × 7 Latin squares," *Annals of Eugenics*, Vol. 14 (1948), p. 133.

[4] A. Sade, *Enumération des carrès latins. Application au 7ème ordre. Conjectures pour les ordres supérieurs*, privately published.

RELATIONS BETWEEN MOMENTS OF ORDER STATISTICS[1]

By Randal H. Cole

University of Western Ontario

Summary. Moments of order statistics multiplied by appropriate factors are called normalized moments. These normalized moments are shown to be successive differences of normalized moments of largest order statistics.

1. Introduction. After discovering the following relations, the author learned in conversation with H. O. Hartley that similar relations had been used by him and others at the University of London. They did not, however, recognize the advantage of expressing them in terms of what are here called "normalized moments." The extreme simplicity of the relations is a direct result of this device.

2. Derivation of the relations. Let $x_{1|n} \geq x_{2|n} \geq \cdots \geq x_{n|n}$ be the order statistics from a sample of size n. Let $g_{i|n}(x)$ be defined by

$$g_{i|n}(x) = F^{n-i}(x)[1 - F(x)]^{i-1}f(x), \tag{1}$$

where $f(x)$ and $F(x)$ are, respectively, the pdf and cdf of the population from which the sample is drawn. Then, the pdf of $x_{i|n}$ is

$$nC_{i-1}^{n-1}g_{i|n}(x),$$

where $C_j^m = m!/[j!(m - j)!]$. If r is any integer such that $r \leq i$, we may write

$$\begin{aligned}[1 - F(x)]^{i-1} &= [1 - F(x)]^{r-1}[1 - F(x)]^{i-r} \\ &= [1 - F(x)]^{r-1} \sum_{j=0}^{i-r} (-1)^{i-r-j}C_j^{i-r}F^{i-r-j}(x).\end{aligned}$$

Substituting this in (1), we have

$$g_{i|n}(x) = \sum_{j=0}^{i-r} (-1)^{i-r-j}C_j^{i-r}g_{r|n-j}(x), \qquad r \leq i \leq n. \tag{2}$$

The relations (2) may be written in matrix form. By letting $C_s^r \equiv 0$ for $s > r$, and considering the expansion of $[1 - (1 - y)]^i$, it can be shown that the inverse of the $(n + 1)$ by $(n + 1)$ matrix

$$P_n = (C_j^i), \qquad i, j = 0, 1, \cdots, n,$$

is

$$P_n^{-1} = ((-1)^{i+j}C_j^i), \qquad i, j = 0, 1, \cdots, n.$$

---

[1] Prepared in connection with research carried out at Princeton University and sponsored by the Office of Naval Research.

We shall introduce the symbol $g'_{p\cdots q|n}(x)$ to represent the vector

$$(g_{p|n}(x), \cdots, g_{q|n}(x)).$$

Similarly let $g'_{i|p\cdots q}(x)$ represent the vector

$$(g_{i|p}(x), \cdots, g_{i|q}(x)).$$

The vector $g_{p\cdots q|n}(x)$ will be the transpose of the vector $g'_{p\cdots q|n}(x)$, etc. A similar notation will be used for vectors whose components are moments.

Using these conventions, the relations (2) can be written

$$g_{r\cdots n|n}(x) = P_{n-r}^{-1} g_{r|n\cdots r}(x), \qquad r < n,$$

or, by inverting,

$$g_{r|n\cdots r}(x) = P_{n-r} g_{r\cdots n|n}(x). \tag{3}$$

Let $\mu_{i|n}$ represent the tth moment of $x_{i|n}$. The omission of t in this designation simplifies the notation and can lead to no confusion, since all the relations between moments are independent of t. Let the quantity

$$\nu_{i|n} = \int_{-\infty}^{\infty} g_{i|n}(x)\, x^t\, dx$$

be called the *normalized tth moment* of $x_{i|n}$. Evidently

$$\mu_{i|n} = nC_{i-1}^{n-1} \nu_{i|n}.$$

If relations (3) are multiplied by x^t and integrated, we have, in terms of the vector notation previously introduced,

$$\nu_{r|n\cdots r} = P_{n-r} \nu_{r\cdots n|n}, \qquad r < n. \tag{4}$$

From this fundamental relation, two special cases of interest can be written down. First, by letting $r = 1$, we have

$$\nu_{1|n\cdots 1} = P_{n-1} \nu_{1\cdots n|n}.$$

Second, because of the triangular nature of P_{n-r}, we may delete the last k components of the two vectors in relation (4), make a corresponding reduction in the order of the matrix, and write

$$\nu_{r|n\cdots r+k} = P_{n-r-k} \nu_{r\cdots n-k|n}.$$

In particular, if $k = n - r - 1$, we obtain

$$\nu_{r|n-1} = \nu_{r|n} + \nu_{r+1|n}.$$

That is to say, the normalized moments for a sample of size $n - 1$ can be obtained by summing adjacent pairs of normalized moments for a sample of size n. It follows that the normalized moments for all samples of size less than n can be obtained, by the simple operation of addition, from the normalized moments

for a sample of size n. By reversing the process, it is clear that if $\nu_{1|n}$, $\nu_{1|n-1}$, $\cdots$, $\nu_{1|1}$ are known, the normalized moments for all samples of size no greater than n can be determined by successive differencing, although in this case there is a progressive loss of significant figures.

CORRECTION TO "THE PROBLEM OF THE GREATER MEAN"

By Raghu Raj Bahadur and Herbert Robbins

University of Chicago and University of North Carolina

In the paper mentioned in the title (*Annals of Mathematical Statistics*, Vol. 21 (1950), pp. 469–487), the paragraph on page 484 beginning "We have given no criterion . . ." is erroneous, and should be omitted. The following paragraph would then read: "Let us suppose that Ω is given by (33). Then $f^0(v)$ is admissible and minimax, by the preceding paragraph. There is, however, another reason for preferring $f^0(v)$ . . .."

We remark that in case a point on the plane $\{\omega : m_1 = m_2\}$ is an interior point of Ω and the risk function is $\bar{r}$, then (contrary to statements in the erroneous paragraph) $f^0(v)$ possesses the following property. *If $f(v)$ is a decision function such that $f(v) \not\equiv f^0(v)$ and*

$$\sup_{\omega\varepsilon\Omega} \bar{r}(f \mid \omega) \leq \sup_{\omega\varepsilon\Omega} \bar{r}(f^0 \mid \omega)(= \tfrac{1}{2}),$$

then $\bar{r}(f^0 \mid \omega) \leq \bar{r}(f \mid \omega)$ for all ω in Ω, the inequality being strict whenever $m_1 \neq m_2$. It follows that $f^0(v)$ is the unique decision function which is admissible and minimax. A proof of this remark is contained in an unpublished paper by R. R. Bahadur entitled "A Property of the t Statistic."

ERRATA

By P. V. Krishna Iyer

University of Oxford

In the author's paper "The theory of probability distributions of points on a lattice" (*Annals of Math. Stat.*, Vol. 21 (1950), pp. 198–217), read "$2 \times 2 \times 3$" for "$2 \times 3 \times 3$" on page 211, line 22, and on page 213, Table 8, heading.

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Oak Ridge meeting of the Institute, March 15–17, 1951)

1. **Confidence Intervals for the Mean Rate at Which Radioactive Particles Impinge on a Type I Counter. (Preliminary Report.)** G. E. Albert, University of Tennessee and Oak Ridge National Laboratory.

The number of particles impinging on a Geiger-Mueller counter in a time interval of length t is assumed to be a random variable with a Poisson distribution of mean at. Starting with Feller's results for a Type I counter given in his paper "On probability problems in the theory of counters" in the *Courant Anniversary Volume*, 1948, it is shown that the count N registered by the counter in time t has the distribution: $Pr(N \geqq m) = 0$ if $m \geqq (t/u) + 1$, and $Pr(N \geqq m) = \exp(-\lambda)\Sigma_{k=m}^{\infty}\lambda^k/k!$, $\lambda = a[t - (m - 1)u]$, if $m < (t/u) + 1$, where u is the dead time of the counter. Confidence interval charts for the parameter $b = au$ for various values of $t' = t/u$ are prepared by the usual inversion procedure. If N and $t-Nu$ are both large, approximate confidence intervals for the parameter a take the simple form

$$(N \pm x_p N^{\frac{1}{2}})/[t - (N - 1)u],$$

where x_p is the two-tailed percentage point of the normal distribution for the confidence level $1 - p$.

2. A Problem of Elapsed Times in a Sequence of Events. Osmer Carpenter, Carbide and Carbon Chemicals Division, Oak Ridge.

The problem considered refers to a series of random events forming a sequence in time or space, for example, the emission of particles by radioactive matter. From a sequence f of such events, a derived sequence g is formed by selecting from f all those events which follow the preceding event by an elapsed time greater than a given constant, $U \geqq 0$. The times between successive events in the sequence f are given to the independently distributed by a known distribution function, $F(t)$. It is required to find the distribution functions of elapsed time and of the number of counts per fixed time interval for the derived sequence, g. A general method is applied to the solution of the exponential case, $F(t) = ke^{-kt}$.

3. On the Existence of Unbiased Tests for Testing Composite Hypotheses. Esther Seiden, University of Buffalo.

The following problem was suggested by J. Neyman. Let X be an observable random variable, multivariate or not, and H a composite hypothesis concerning X. Let $\bar{H}$ denote a hypothesis, concerning X, alternative to H. Finally, let α be a chosen level of significance. What restriction should one impose on the hypotheses H and $\bar{H}$ in order that there exists a critical region w such that (i) $P(X \varepsilon w \mid H) = \alpha$, and, whatever be the simple hypothesis $h \varepsilon \bar{H}$, (ii) $P(X \varepsilon w \mid h) > \alpha$? It is shown now that if H as well as $\bar{H}$ consists in assuming that the random variable X follows a continuous distribution law, then there exists always the most powerful region w satisfying conditions (i) and (ii), provided that the distributions belonging to H and $\bar{H}$ are linearly independent. If H and $\bar{H}$ are infinite families of absolutely continuous distributions and condition (i) is replaced by (i') $P(X \varepsilon w \mid H) \leqq \alpha$, then for some α less than $\frac{1}{2}$ there exists a region w satisfying conditions (i') and (ii), provided that the convex closures of H and $\bar{H}$ are disjoint.

4. Group Divisible Incomplete Block Designs. R. C. Bose, University of North Carolina.

An incomplete block design with v treatments each replicated r times in b blocks of size k is said to be group divisible if the treatments can be divided into m groups each with n treatments, so that the treatments of the same group occur together in λ_1 blocks and treatments of different groups occur together in λ_2 blocks, $\lambda_1 \neq \lambda_2$. The parameters are connected by the relations $v = mn$, $bk = vr$, $\lambda_1(n - 1) + \lambda_2 n(m - 1) = r(k - 1)$. It is shown that these designs fall into three classes: (i) singular for which $r = \lambda_1$, (ii) semiregular for which $r > \lambda_1$, $rk = v\lambda_2$, (iii) regular for which $r > \lambda_1$, $rk > v\lambda_2$. It is proved that for regular designs $b \geqq v$, and for semiregular designs $b \geqq v - m + 1$, every block containing

the same number of treatments from each group. A singular design is always derivable from a balanced incomplete block design by replacing each treatment by a group of n new treatments. When $b = v$ the quantity $(r - \lambda_1)^{m(n-1)}(rk - v\lambda_2)^{m-1}$ must be a perfect square, and the Hasse invariant of NN', where N is the incidence matrix of the design, must be $+1$. The value of this invariant has been calculated in terms of the parameters. The parameters for all group divisible designs with $r \leqq 10$, $k \leqq 10$, whose existence is not ruled out by theorems stated above, have been listed. Combinatorial solutions for most of these have been derived, though there remain a number of unsolved cases. The analysis of variance and the equations for intra- and inter-block estimates have been given. These designs are likely to prove useful both in varietal trials and in factorial experiments.

5. **Orthogonal Arrays of Strength Two and Three.** R. C. Bose and Kenneth A. Bush, University of North Carolina.

Consider a matrix $A = ((a_{ij}))$ with m rows and N columns where each element a_{ij} represents one of the s integers $0, 1, 2, \cdots, s - 1$. The columns of any t-rowed submatrix of A provide N ordered t-plets. The matrix A is called an orthogonal array (N, m, s, t) of size N, m constraints, s levels, and strength t if each of the C_t^m partial t-rowed matrices formed from A contains all the s^t possible ordered t-plets each repeated λ times $(N = \lambda s^t)$. The known upper bounds for the number of constraints when $t = 2$ and 3 have been improved: If $\lambda - 1 = a(s - 1) + b$, $0 \leqq b < s - 1$, and n is the largest positive integer (including 0) consistent with $s(b - 2n) \geqq (b - n)(b - n + 1)$, then for the case $t = 2$, $m \leqq I[(\lambda s^2 - 1)/(s - 1)] - n - 1$, and for the case $t = 3$, $m \leqq I[(\lambda s^2 + s - 2)/(s - 1)] - n - 1$. Methods of constructing orthogonal arrays of strength 2 and 3 have been investigated. A difference theorem enabling the construction of the arrays (18, 7, 3, 2) and (32, 9, 4, 2) has been proved, and it is shown that if $s = p^n$, $\lambda = p^u$, where p is a prime, then we can construct the array $(\lambda s^2, m, s, 2)$, with

$$m = 1 + p^{n+d} + \cdots + p^{rn+d} + p^{rn+n+d},$$

where $u = rn + d$, $0 \leqq d < n$, $r \geqq 0$. Another theorem connects finite projective geometries with orthogonal arrays, and is used to construct the arrays (i) $(s^3, s + 2, s, 3)$ when $s = 2^n$; (ii) $(s^3, s + 1, s, 3)$ when $s = p^n$, p being an odd prime; (iii) $s^4, s^2 + 1, s, 3)$ when $s = p^n$, where p is a prime; (iv) $(s^r, s^{r-1}, s, 3)$ when $s = 2$. Orthogonal arrays are useful in connection with many problems of experimental design.

6. **The Structure of Balanced Incomplete Block Designs, and the Impossibility of Certain Unsymmetrical Cases.** William S. Connor, University of North Carolina.

If α_{ij} is the number of treatments common to the ith and jth blocks of a balanced incomplete block design with v treatments, b blocks, r replications, and k treatments per block, with any two treatments occurring together in the same block λ times, then the characteristic matrix $C = ((c_{ij}))$ of the design may be defined by $c_{ii} = (r - k)(r - \lambda)$, $i = 1, 2, \cdots, v$, $c_{ij} = k\lambda - r\alpha_{ij}$, $i, j = 1, 2, \cdots, v$, $i \neq j$. If $|C_t|$ is any symmetrically chosen partial determinant of order r belonging to C, we prove that (i) $|C_t|$ is nonnegative; (ii) if $t = b - v$, then $|C_t| k(r - \lambda)^{v-t-1}/r^{t-1}$ is a perfect square; (iii) if $t > b - v$, $|C_t| = 0$. From (i) Fisher's inequality $b \geqq v$ is deduced and it is shown that

$$k + \lambda - r \leqq \alpha_{ij} \leqq r - \lambda - k + 2k\lambda/r.$$

The structure of the designs (a) $v = 15$, $b = 21$, $r = 7$, $k = 5$, $\lambda = 2$; (b) $v = 36$, $b = 45$, $r = 10$, $k = 8$, $\lambda = 2$; (c) $v = 21$, $b = 28$, $r = 8$, $k = 6$, $\lambda = 2$ is studied. For (a) and (b) it is proved that there must exist $b - v$ blocks, the $|C_{b-v}|$ for which contradicts (ii). For

(c) it is shown that if the incidence matrix N is augmented to N_0 by adding 7 suitably chosen row vectors then the Hasse invariant $C_p(N_0N_0')$ for N_0N_0' is -1, when $p = 3$. This demonstrates the impossibility of (a), (b), and (c). The last two results are new. (Research carried on under the sponsorship of the Office of Naval Research.)

7. Some Bounded Significance Level Tests for the Median. JOHN E. WALSH, The Rand Corporation.

In practice it is often permissible to assume that the observations of a set are statistically independent and from continuous populations with a common median. This is the case, for example, if the observations are a sample from a continuous population. Then the population median can be investigated by using the sign test. For small numbers of observations, however, the sign test does not furnish very many suitable significance levels. Also, some of the sign tests with suitable significance levels are not very efficient. This note presents some tests whose significance levels are only approximate but cover a wide range of suitable values. The significance levels of these tests are exactly determined if the populations are also symmetrical; they are bounded otherwise. Some of these bounded significance level tests have high efficiencies.

8. Joint Sampling Distribution of the Mean and Standard Deviation for Distribution Functions of the First Kind. MELVIN D. SPRINGER, U. S. Naval Ordnance, Indianapolis.

Consider a universe characterized by the distribution function $f(x)$, $-\infty < x < \infty$. If n variates x_i, $i = 1, 2, \cdots, n$, are selected at random from this universe, the probability that they will fall simultaneously within the intervals dx_i, $i = 1, 2, \cdots, n$, is given, to within infinitesimals of higher order, by $f(x_1)f(x_2) \cdots f(x_n)\, dx_1\, dx_2 \cdots dx_n$. As an immediate consequence of the definitions of $\bar{x}$ and s one may employ the transformation T: $x_1 = x_1$, $x_2 = x_2$, $\cdots$, $x_{n-2} = x_{n-2}$, $x_{n-1} = (n\bar{x} - \Sigma_1^{n-2} x_i \pm \Omega_1)/n$, $x_n = n\bar{x} - \Sigma_1^{n-1} x_i$, where $\Omega_1 = [-3\Sigma_1^{n-2} x_i^2 - 2\Sigma_{i=1}^{n-3}\Sigma_{j=i}^{n-3} x_i x_{j+1} + 2n\bar{x}\,\Sigma_1^{n-2} x_i - n(n-2)\bar{x}^2 + 2ns^2]^{1/2}$. Application of this transformation gives

$$f(x_1)f(x_2) \cdots f(x_n)\, dx_1\, dx_2 \cdots dx_n$$
$$= f(x_1)f(x_2) \cdots f(x_{n-2})f([n\bar{x} - \Sigma_1^{n-2} x_i - \Omega_1]/2)$$
$$\cdot f([n\bar{x} - \Sigma_1^{n-2} x_i + \Omega_1]/2) \mid J \mid dx_1\, dx_2 \cdots dx_{n-2}\, d\bar{x}\, ds,$$

where $\mid J \mid = \mid$ Jacobian of $T \mid = n^2 s/\Omega_1$. Evaluation of the multiple integral

$$F(\bar{x}, s) = \int\int \cdots \int f(x_1)f(x_2) \cdots f(x_{n-2})f([n\bar{x} - \Sigma_1^{n-2} x_i - \Omega_1]/2)$$
$$\cdot f([n\bar{x} - \Sigma_1^{n-2} x_i + \Omega_1]/2)\; 2n^2 s/\Omega_1\, dx_{n-2} \cdots dx_2\, dx_1$$

yields the joint distribution function $F(\bar{x}, s)$. The limits of integration are established by employing the relationships $\Sigma_1^n x_i = n\bar{x}$ and $\Sigma_1^n x_i^2 = ns^2 + n\bar{x}^2$, together with mathematical induction, to prove that x_{n-r}, $r = 2, 3, \cdots, n-1$, is restricted to the closed interval

$$([n\bar{x} - \Sigma_1^{n-r-1} x_i - \Omega_r]/(r+1),\ [n\bar{x} - \Sigma_1^{n-r-1} x_i + \Omega_r]/[r+1]),$$

where

$$\Omega_r = [-r(r+2)\Sigma_1^{n-r-1} x_i^2 - 2r\Sigma_{i=1}^{n-r-2}\Sigma_{j=i}^{n-r-2} x_i x_{j+1}$$
$$+ 2rn\bar{x}\Sigma_1^{n-r-1} x_i - rn(n-r-1)\bar{x}^2 + (r+1)rns^2]^{1/2}.$$

9. On Certain Distribution Problems in Multivariate Analysis. (Preliminary Report.) INGRAM OLKIN, University of North Carolina.

This paper is concerned with the derivation of the joint distributions of (i) rectangular coordinates, (ii) correlation coefficients, (iii) characteristic roots of a matrix, and (iv) roots of a determinantal equation, starting in each case from the multivariate normal distribution. Consider a set of pn random variables following the distribution law $f(X, n) = K \exp(-\frac{1}{2} tr\, XX')$, X a $p \times n$ matrix, and the real transformations: (1) $X = (T\,O)e^{A}$, where T $(p \times p)$ is a triangular matrix with $t_{ij} = 0$ $(i < j)$, A $(n \times n)$ is a skew-symmetric matrix; (2) $X = D_\alpha(U\,O)e^{B}$, $UU' = R$, where D_α is a diagonal matrix with elements $\alpha_1, \cdots, \alpha_p$, U $(p \times p)$ is a triangular matrix with $u_{ij} = 0$ $(i < j)$, $\Sigma_j^p u_{ij}^2 = 1$, $i = 1, \cdots, p$, B $(n \times n)$ is a skew-symmetric matrix, R $(p \times p)$ is a symmetric matrix; (3) $X = e^{C}(D_\mu\, O)e^{D}$, where C $(p \times p)$ and D $(n \times n)$ are skew-symmetric, D_μ is a diagonal matrix with elements $\mu_1, \cdots, \mu_p$, where μ^2 are the characteristic roots of XX'. Using these transformations on $f(X, n)$, (i), (ii), and (iii) are obtained. From the distribution law

$$f(X_1, n_1)f(X_2, n_2)$$

and the transformation (4) $X_1 = Y(D_s O)e^{E}$, $X_2 = Y(D_c O)e^{F}$, where Y $(p \times p)$, E $(n_1 \times n_1)$ and F $(n_2 \times n_2)$ are skew-symmetric matrices, D_s and D_c are diagonal matrices with elements $s_1, \cdots, s_p$ and $c_1, \ldots, c_p$ respectively, such that $s^2 + c^2 = 1$, the joint distribution of $\theta = s^2$ is found, where θ are the roots of $| X_1X_1' - \theta(X_1X_1' + X_2X_2') | = 0$.

10. A Unified Approach to a Wide Class of Distribution Problems in Multivariate Analysis. S. N. ROY, University of North Carolina.

(1) X being a $p \times n$ matrix of random observations (reduced to means) from a p-variate normal population, and it being known that there exist a $p \times p$ triangular matrix T, and a $p \times n$ matrix L (both ordinarily uniquely determined) such that $X = TL$ and $LL' = I$, it is of interest to obtain the sampling distribution of T from which various distributions, including those of partial and multiple correlations, would easily follow. (2) X_1 and X_2 being $p \times n_1$ and $p \times n_2$ matrices of random observations (reduced to means) from two p-variate normal populations and it being known that there exist (ordinarily uniquely) a $p \times p$ matrix Z, and a $p \times n_1$ matrix L_1, a $p \times n_2$ matrix L_2 (with the constraints $L_1L_1' = L_2L_2' = I$), and $p \times p$ diagonal matrices D_s and D_c (where $S_i = \text{Sin}\ \theta_i$; $C_i = \text{Cos}\ \theta_i$; $i = 1, 2, \cdots, p$) such that $X_1 = ZD_sL_1$, $X_2 = ZD_cL_2$, it is of interest in multivariate analysis to obtain the sampling distribution of θ $(\equiv \theta_1, \cdots, \theta_p)$. (3) With the same X matrix as in (1), and it being known that there exist (ordinarily uniquely) a $p \times p$ orthogonal matrix Γ, a $p \times n$ matrix M (such that $MM' = I$), and a diagonal matrix $D_t(p \times p)$ such that $X = \Gamma D_t M$, it is of interest to obtain the sampling distribution of $t(\equiv t_1, t_2, \cdots, t_p)$. With the help of the constraints indicated one could knock out any $p(p + 1)/2$ out of L in (1), out of each of L_1 and L_2 in (2), and of each of Γ and M in (3); denote the remaining elements respectively by L_R, $(L_{1R}, L_{2R}($, and (Γ_R, M_R). Then in (1), (2), and (3), respectively, we change over from X to (T, L_R), from (X_1, X_2) to $(Z, \theta, L_{1R}, L_{2R})$, and from X to (Γ_R, t, M_R). This is made easy by an artifice discussed in the paper, and the way L in (1) (L_1, L_2) in (2), and M in (3) occur, makes it easy to integrate out over them leaving us with the distributions of T in (1), (Z and θ) in (2), and (Γ and t) in (3). From this the null distributions of T in (1), θ in (2), and t in (3) follow with great ease. Certain nonnull distributions would also come out without much difficulty.

11. An Extension of the Buffon Needle Problem. NATHAN MANTEL, National Cancer Institute.

Historically, the Buffon needle problem is concerned with the estimation of the value of π from the probability of intersection of a straight line of fixed length (< 1) with a series

of equally unit-spaced parallel lines, on which the straight line is allowed to fall at random. The present paper extends the problem to the estimation of π from the average number of intersections of a straight line of any fixed length with a series of equally spaced parallel and perpendicular lines on which the straight line is allowed to fall at random. It is also shown that, comparatively, very precise estimates of π can be made, for long straight lines, from the variation in number of intersections rather than from the average number of intersections. From purely statistical considerations it is demonstrated that π must lie between 3.1231 and 3.1752, with no necessity for any measurements being made.

12. A Generalization of Sampling without Replacement from a Finite Universe. D. G. Horvitz and D. J. Thompson, Iowa State College.

Let the finite universe consist of N elements U_i $(i = 1, 2, \cdots, N)$. A sample of n elements is to be drawn without replacement and the total T of some character X of the elements estimated from the sample. Denote by $P(U_i)$ the probability that the ith element will be included in a sample of size n. An unbiased estimator $\hat{T} = \sum_{i=1}^{n} x_i/P(U_i)$ is proposed, and expressions for the variance of this estimator as well as an unbiased estimator of this variance are given. An extension to a two-stage sampling scheme is presented. Consideration is given to methods of determining selection probabilities which will result in optimum probabilities $P(U_i)$ on the basis of the prior information available on the elements of the universe, and two approximate methods are illustrated.

13. A Problem in Two-Stage Sampling. B. M. Seelbinder, University of North Carolina.

Charles Stein has suggested a two-stage sampling plan, the size of the second part of the sample depending on the information supplied about the variance of the population by the first part of the sample. In his work, the size n_1 of the first part of the sample is left to the discretion of the experimenter. This study is designed to throw further light on the choice of the value of n_1. For this purpose the expected value of the total sample size n for given n_1 has been computed for four different significance levels $\alpha = .1, .05, .02, .01$ and varying $c = d/\sigma$, where d is the allowable discrepancy. These values are presented in four tables where c ranges from .01 to 1.0 and n_1 ranges from 5 to 72,000. It is shown that the computation can be made to depend on the knowledge of Pearson's incomplete Gamma function. An approximation whereby the computation can be made to depend only on the knowledge of the normal distribution function has also been developed. Numerical evidence for the adequacy of the approximation for moderately large values of n_1 $(n_1 \geqq 61)$ has been adduced. Limiting values for the expected value of the total sample size are given for fixed n_1 and α with varying c. The discussion of the use of the tables covers the different sampling situations which may arise: (i) an approximate estimate of σ is available, (ii) only a rough estimate of σ is available. Reasons are given which point to 250 as the upper limit for n_1 in a two-stage sampling plan.

14. Bounds on a Distribution Which Are a Function of Moments to Order Four. (Preliminary Report.) Marvin Zelen, University of North Carolina.

Let $F(y)$ be a cumulative distribution function defined for the random variable $a < y < b$, and x be a known quantity. Markov and Stieltjes considered the problem of finding the $\inf_{F(y)} F(x)$ and the $\sup_{F(y)} F(x)$ as a function of a finite number of moments of the distribution. This present paper investigates the explicit expressions for these bounds if the moments to order four are known (1) in the case when the random variable has finite range, (2) in the case when the random variable has infinite range. In the applications of these bounds, it is necessary to order roots of certain orthogonal polynomials. It is suggested that for ready application, a nomograph be used. These bounds would be useful when one

is confronted with a cumulative distribution function which is unknown or difficult to handle.

15. An Inconsistency among Type A Regions. Herman Chernoff, University of Illinois.

In a test of a hypothesis one may regard a sample in the critical region as evidence that the hypothesis is false. Let us assume that for some reason it is desired to increase the critical size of the test, i.e., to make rejection of the hypothesis more probable. Then one may expect that an observation which led to rejection in the first test should still lead to rejection in the new test. In other words, one should expect $W_\alpha \supset W_{\alpha'}$ if $\alpha > \alpha'$ where W_α is the critical region of size α. An example is given where regions of Type A fail to have this property.

16. Stochastic Approximation. (Preliminary Report.) Herbert Robbins and S. Monro, University of North Carolina.

We consider the general problem of estimating a constant associated with a function $M(x)$, e.g., the root of an equation $M(x) = \alpha$ or the abscissa of a maximum of $M(x)$. When $M(x)$ is observable there are methods of determining the constant by "successive approximation." We suppose, on the contrary, that $M(x)$ is unknown but that to each value of x corresponds an observable random variable $Y = Y(x)$ with distribution function $P[Y(x) \leqq y] = H(y \mid x)$ such that $M(x) = \int y\, dH(y \mid x)$ is the expected value of Y for the given x. In the case where $M(x)$ is increasing and we wish to estimate the unique root, $x = \theta$, of $M(x) = \alpha$, we propose to let $x_{n+1} = x_n + a_n(\alpha - y_n)$, where x_1 is an arbitrary constant, $\{a_n\}$ is a sequence of positive constants, and y_n is a random variable such that $P[y_n \leqq y \mid x_n] = H(y \mid x_n)$. One of us has shown under certain conditions on $H(y \mid x)$ that, if $\{a_n\}$ is of the type $\{1/n\}$, then no matter what the initial value x_1,

$$\lim_{n\to\infty} E(x_n - \theta)^2 = 0,$$

so that x_n is a consistent estimator of θ. Work is in progress to establish in special cases bounds on $E(x_n - \theta)^2$. By replacing $(\alpha - y_n)$ by $\operatorname{sgn}(\alpha - y_n)$, less severe restrictions need be imposed on $H(y \mid x)$ in order to obtain the same result. This sequential type of design can be applied to estimation in regression problems, to "all or nothing" response experiments (where y_n is limited to the values 0 and 1), and to the experimental determination of the maximum of a function (cf. Harold Hotelling's paper in *Annals of Mathematical Statistics*, vol. 12 (1941), pp. 20–45). (This research was done in part under an Office of Naval Research contract.)

17. On the Properties and Statistical Purposes of Some Well-known and Some New Tests in Multivariate Analysis. S. N. Roy, University of North Carolina.

Consider two problems of multivariate analysis, each of which could be made to cover a wide number of situations. (1) With two random samples of sizes n_1 and n_2 and dispersion matrices (a_{1ij}) and (a_{2ij}) from two p-variate normal populations with dispersion matrices (α_{1ij}) and (α_{2ij}), $i, j = 1, 2, \cdots, p$, an infinite number of similar region tests could be constructed for the composite hypothesis $(\alpha_{1ij}) = (\alpha_{2ij})$, among which there is none having the strong optimum properties of the usual F-test in the anlogous univariate problem. Among these similar region tests, however, there is a subset based on F_i, $i = 1, 2, \cdots, p$,

where F_i's are the roots of the equation in $F: |a_{1ij} - Fa_{2ij}| = 0$, such that the largest root has moderate optimum properties with respect to one class of alternatives, the smallest for another class and the product of the roots (which is the likelihood ratio test) for another class of alternatives—all discussed in this paper. (2) With k random samples of sizes n_r from k p-variate normal populations with means m_{ri} and a common dispersion matrix (α_{ij}), $r = 1, 2, \cdots, k$; $i, j = 1, 2, \cdots, p$, an infinite number of similar region tests could be constructed for the composite hypothesis $(m_{1i}) = (m_{2i}) = \cdots = (m_{ki})$, $i = 1, 2, \cdots, p$, among which there is none having the strong optimum properties of the F-test in the analogous univariate problem. Among these similar region tests, however, there is a subset based on F_i, $i = 1, 2, \cdots, q \leq p$, where F_i's are the nontrivial roots of the equations in F: $| b_{1ij} - Fb_{2ij} | = 0$ (where (b_{1ij}) is the matrix of the sample means reduced to the grand means and (b_{2ij}) is the pooled dispersion matrix from the different samples), such that the largest and smallest roots have moderate optimum properties with respect to two different classes of alternatives and the sum of the roots for a third class of alternatives—all discussed in this paper. The likelihood ratio test, however, leads to the product. The wide variety of situations each problem could be made to cover is also discussed in this paper.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest.

Personal Items

Dr. K. S. Banerjee, Statistician at the Central Sugarcane Research Station, Bihar, India, received his doctorate degree from the Calcutta University in January of this year. His thesis covered his contributions to "weighing designs."

Mr. Lyle D. Calvin, formerly at the Institute of Statistics, North Carolina State College, has accepted the position of Biometrician with the Division of Biological Research, G. D. Searle & Co., Chicago, Illinois.

Dr. Robert J. Hader has accepted a position on the staff of the Institute of Statistics, North Carolina State College. He leaves Los Alamos, New Mexico, where he has been employed as statistician for the Los Alamos Scientific Laboratory for the past two years.

Mr. Bernard Hecht has joined the Victor Division of RCA, Camden, New Jersey, as Manager, Assembly Quality Control, after five years as Quality Control Manager of the International Resistance Company of Philadelphia, Pennsylvania.

Dr. Edward L. Kaplan has received his doctorate degree in mathematics from Princeton University and is now a member of the Technical Staff, Bell Telephone Laboratories, Murray Hill, New Jersey.

Dr. Eugene Lukacs has joined the staff of the Statistical Engineering Laboratory of the National Bureau of Standards. At the Bureau he will be engaged in research in mathematical statistics, particularly autoregressive series and stochastic processes.

Mr. A. W. Marshall, formerly at the Washington, D. C., office of the Rand Corporation, has now moved to its Santa Monica, California, office.

Dr. J. E. Morton is on leave of absence from Cornell University and is serving as Chief, Statistical Research and Development Staff, Office of the Housing Administrator, Washington, D. C.

Dr. J. Ernest Wilkins, Jr., formerly on the staff at the American Optical Company, is now with Nuclear Development Associates, Inc., 80 Grand Street, White Plains, New York.

Dr. William J. Youden of the National Bureau of Standards has recently been elected to Fellowship in the New York Academy of Sciences.

Sigeiti Moriguti, Assistant Professor of Applied Mathematics at the University of Tokyo, is spending the academic year 1950–51 in research and study of mathematical statistics at the University of North Carolina under the sponsorship of the United States Army. He is the author of numerous research articles and a book on the theory of statistics.

Statistics at Chicago

The University of Chicago in 1949 established a Committee on Statistics which is in all respects equivalent to a department, having its own faculty, budget, and curriculum. Its purposes are research, instruction, and consultation. Its faculty includes R. R. Bahadur, Milton Friedman, Leo A. Goodman, John Gurland, Tjalling C. Koopmans, William H. Kruskal, Harry V. Roberts, Murray Rosenblatt, Leonard J. Savage, Charles M. Stein, and W. Allen Wallis (Chairman). Among other statisticians at the University of Chicago are Walter Bartky (Physical Sciences), Donald W. Fiske (Psychology), Philip M. Hauser (Sociology), Paul R. Halmos (Mathematics), Karl J. Holzinger (Education), H. Gregg Lewis (Economics), Jacob Marschak (Cowles Commission for Research in Econometrics), William Stephenson (Psychology), Louis L. Thurstone (Psychology), Josephine Williams (National Opinion Research Center), and Sewall Wright (Zoology).

The following courses are offered by the Committee: Introduction to Statistics; Statistical Inference (3 quarters); Introduction to Mathematical Probability; Introduction to Mathematical Statistics (2 quarters); Sample Surveys; Analysis of Variance and Regression; Estimation and Tests of Hypotheses; Statistical Theory of Decision-making; Theory of Minimum Risk; Sequential Analysis; Non-parametric Inference; Multivariate Analysis; Design of Experiments; Time Series; Statistical Problems of Model Construction; Limit Theorems of Probability Theory; Markov Processes; Mathematical Techniques of Statistics; and several seminars. In addition, a number of statistics courses are offered in other departments, e.g., Factor Analysis, Econometrics, Quality Control, Index Numbers, Biometrics, etc.

Three kinds of degree may be obtained at Chicago in Statistics. (1) *The M.A. or Ph.D. in a substantive field, with concentration in Statistics*, is not administered by the Committee, but it cooperates fully with the substantive departments in these degree programs. (2) *The M.A. in Statistics* is awarded on the basis of

(i) a thesis, (ii) written examinations, and (iii) work in a minor field. (3) *The Ph.D. in Statistics* is awarded on the basis of (i) preliminary written examinations, (ii) work in a minor field, (iii) participation in statistical consultation, (iv) a dissertation, (v) a public lecture on the content of the dissertation, and (vi) a final oral examination.

Summer Sessions in Berkeley, California

This year's summer program at the Statistical Laboratory of the University of California, Berkeley, California, consists of two sessions, June 18–July 28 and July 30–September 8. The program includes four of the usual undergraduate courses, two in each session, and two graduate courses. One of the latter is a regular course of lectures on rank correlation methods and on time series analysis. The other graduate course is a seminar on time series and related problems. Both graduate courses will be given during the first summer session by Professor Maurice G. Kendall of the London School of Economics and Political Science. Professor J. Neyman will be available for consultations on work leading to higher degrees. In addition to the above two persons, the faculty of the summer sessions will include Dr. Grace E. Bates (Mount Holyoke College), Dr. Colin R. Blyth (University of Illinois), and Dr. Gottfried E. Noether (New York University).

Summer Seminar in Statistics

The second annual session of the Summer Seminar in Statistics will be held at the University of Connecticut, Storrs, Connecticut, August 6–31, 1951. The purpose of the Seminar is to stimulate general exchange of ideas by providing informal contacts and free discussions among academic statisticians, students, and users of statistical techniques. The principal session meets daily from 3:00 p.m. to supper. The schedule of topics, together with the organizers of each week's program, is as follows:

August 6–10.	Statistics in the Biological Sciences (C. I. Bliss and J. Ipsen);
August 13–17.	Time Series (M. G. Kendall and J. W. Tukey);
August 20–24.	Statistical Theory and Probability (M. Kac and H. Robbins);
August 27–31.	Statistical Techniques with Special Reference to the Social Sciences (F. C. Mosteller, F. L. Strodtbeck, and M. A. Woodbury).

Frequent statistical clinics to discuss the solution of particular practical problems are planned.

Dormitory accommodations of single and double rooms are available at the University of Connecticut. Family groups of three or more must use other housing. It is hoped that a number of stipends to cover living expenses will be available on a competition basis to graduate students. Further information about the Seminar may be obtained from the Secretary of the Seminar: Professor D. F. Votaw, Jr., Department of Mathematics, Yale University, New Haven, Connecticut.

Doctoral Dissertations in Statistics, 1950

Listed below are the doctorates conferred during the year 1950 in the United States and Canada for which the dissertations were written on topics in statistics. The university, month in which degree was conferred, major subject, minor subject, and the title of the dissertation are given in each case if available. If any doctorate properly belonging in this list is omitted, the Editor would like the relevant information concerning such doctorate. It is planned to publish a list of doctorates in the June issue each year.

R. R. Bahadur, North Carolina, June, major in mathematical statistics, minor in experimental statistics and mathematics, "On a Class of Decision Problems in the Theory of *R* Populations."

C. R. Blyth, California, June, major in mathematics, "I. Contribution to the Statistical Theory of the Geiger-Müller Counter. II. On Minimax Statistical Decision Procedures and Their Admissibility."

K. A. Bush, North Carolina, August, major in mathematical statistics, minor in mathematics and economics, "Orthogonal Arrays."

A. L. Finkner, North Carolina, major in experimental statistics, minor in agronomy, "Further Investigation on the Theory and Application of Sampling for Scarcity Items."

W. D. Foster, North Carolina, major in experimental statistics, minor in meteorology, "On the Selection of Predictors: Two Approaches."

M. Halperin, North Carolina, August, major in mathematical statistics, minor in experimental statistics and mathematics, "Estimation in Truncated Sampling Processes."

H. M. Hughes, California, September, major in mathematics, "Estimation of the Variance of the Bivariate Normal Distribution."

P. E. Irick, Purdue, February, major in mathematics, minor in psychology, "A Geometric Study of the Exact Sampling Distribution of Standard Deviations When the Sampled Population Is Arbitrary."

S. L. Isaacson, Columbia, June, major in mathematical statistics, minor in mathematics, "On the Theory of Unbiased Tests of Simple Statistical Hypotheses Specifying the Values of Two or More Parameters."

E. H. Jebe, North Carolina, major in experimental statistics, minor in agricultural economics, "The Theory and Application of the Selection of Primary Units for Sampling an Agricultural Population."

A. W. Kimball, Jr., North Carolina, major in experimental statistics, minor in mathematics, "Studies in the Statistical Design and Analysis of Microbiological Assays of Amino Acids."

G. E. McCreary, Iowa State College, June, major in statistics, minor in mathematics and economics, "Cost Functions for Sample Surveys."

L. E. Moses, Stanford, major in statistics, minor in mathematics, "An Iterative Construction of the Optimum Sequential Procedure When the Cost Function Is Linear."

S. W. Nash, California, June, major in mathematics, "I. Contribution to the

Theory of Experiments with Many Treatments. II. On the Law of the Iterated Logarithm for Dependent Random Variables."

R. P. Peterson, California (Los Angeles), June, major in mathematics, "Certain Optimum Statistical Decision Methods."

M. Pizzi (Doctor of Public Health), Johns Hopkins, "An Approximate Solution for the Standard Error of LD50 as Obtained by the Reed-Muench Method."

B. Sherman, Princeton, June, major in mathematics, "A Random Variable Related to the Spacing of Sample Values."

S. S. Shrikhande, North Carolina, August, major in mathematical statistics, minor in experimental statistics, "Construction of Partially Balanced Designs and Related Problems."

H. Solomon, Stanford, major in statistics, "Distribution of the Measure of a Two-Dimensional Random Set."

H. E. Teicher, Columbia, June, major in mathematical statistics, minor in mathematics, "On the Factorization of Distributions."

W. A. Vezeau, St. Louis, June, major in mathematics, "On the Product Distribution of Normally Distributed Variables."

S. A. Vora, North Carolina, June, major in mathematical statistics, minor in experimental statistics, "Bounds on the Distribution of Chi-Square."

J. T. Wakeley, North Carolina, major in experimental statistics, minor in meteorology, "On Linear Regression Method as Related to Long Time Experiments in Agricultural Climatology."

New Members

The following persons have been elected to membership in the Institute.

(December 1, 1950 to February 28, 1951)

Benktander, Gunnar, Fil. Kand. (Univ. of Stockholm), Actuary, Post Fack, Stockholm 26, Sweden.

Boll, C. H., B.S. (Stanford Univ.), Graduate student in Statistics, Stanford University, *1247 Cowper, Palo Alto, California.*

Carey, T. M., Ph.D. (Univ. of London), Lecturer in Mathematics, University College, Cork, Ireland, *"Duinin," Laburnum Park, Model Farm Road, Cork, Ireland.*

DeLancie, R. H., A.B. (Univ. of Calif.), Graduate student in Statistics, University of California, *1137 Colusa Avenue, Berkeley 7, California.*

Dighero, Oscar Alfonso Martinez, Graduate (Univ. of San Carlos, Guatemala), Civil Engineer, Chief, Division of Engineering and Architecture, Instituto Gautemalteco de Seguridad Social, *11 Avenida Norte No. 44, Guatemala, Guatemala, Central America.*

Ellery, J. B., M.A. (Univ. of Colo.), Graduate student and Teaching Assistant, Department of Speech, University of Wisconsin, *7 Tilton Terrace, Madison 4, Wisconsin.*

Esary, J. D., A.B. (Whitman College), Teaching Assistant, Statistical Laboratory, University of California, *2534 Dwight Way, Berkeley 4, California.*

Esscher, Fredrik, Ph.D. (Univ. of Lund), Chief Actuary, Skandia Insurance Co., Stockholm 2, Sweden.

Grometstein, A. A., M.A. (Columbia Univ.), Industrial Statistician and Consulting Physicist, Sylvania Electric Products, Inc., 70 Forsyth Street, Boston, Massachusetts.

Gross, F. A., B.S. (American Univ.), Statistician, Research Division, Bureau of Naval Personnel, Arlington, Virginia, *3409B New Mexico Ave., N. W., Washington 16, D. C.*

Houthakker, H. S., Ph.D. (Univ. of Amsterdam), Research Officer, University of Cambridge, Department of Applied Economics, *8 Richmond Road, Cambridge, England.*

Hubbell, C. H., A.B. (Oberlin College), Box 467, Benjamin Franklin Station, Washington 4, D. C.

Kurtz, T. E., A.B. (Knox College), Research Assistant, Mathematics Department, Fine Hall, Box 708, Princeton University, Princeton, New Jersey.

Lanteli, Gunnar, Fil. Kand. (Univ. of Lund), Actuary of Försäkringsaktiebolaget Hansa, Stockholm 7, Sweden.

Matern, Bertil, Fil. Lic. (Univ. of Stockholm), Assistant Professor, Swedish Forest Research Institute; *Lappkarrsvagen 47, Stockholm 50, Sweden.*

McCall, Jr., C. H., A.B. (George Washington Univ.), Assistant in Statistics, *6701–44th Street, Chevy Chase 15, Maryland.*

McHugh, R. B., M.A. (Univ. of Minn.), Assistant Professor of Psychology and Statistics, Iowa State College, *222 Stanton, Ames, Iowa.*

Meier, Paul, M.A. (Princeton Univ.), Statistician, and Research Secretary, Philadelphia Tuberculosis and Health Association, *39 Vandeventer Avenue, Princeton, New Jersey.*

Owen, D. B., M.S. (Univ. of Wash.), Research Associate in Mathematics, University of Washington, *612 West 85th Street, Seattle 7, Washington.*

Poch, F. A., Licenciado en Ciencias (Univ. of Madrid), Official, Instituto Nacional de Estadistica; Specialist of Section of Methodology; Assistant Professor, Mathematical Statistics, University of Madrid, Madrid, Spain.

Rios, Sixto, Ph.D., Professor of Mathematical Statistics, University of Madrid; Chief, Department of Statistics of the Superior Council of Scientific Research, Madrid, Spain.

Rosenbaum, S. Z., B.S. (Univ. of Chicago), Research Director, Community Welfare Council of Milwaukee County, 2965 N. 81st Street, Milwaukee 10, Wisconsin.

Saxen, Tryggwe, B.Sc. (Univ. of Helsingfors), Fil. Kand., Assistant Actuary of Industrial Accident Insurance Co., *Kasarngatan 44V, Helsingfors, Finland.*

Shapiro, Arthur, A.B. (Brooklyn College), 6633½ Telegraph Avenue, Oakland 9, California.

Shimizy, Kunio, B.A. (Univ. of British Columbia), Statistician, Institutions Section, Dominion Bureau of Statistics, Ottawa, Ontario, Canada.

Soda, Takemune, M.P.H. (Johns Hokpins Univ.), Chief, Division of Health and Welfare Statistics, Ministry of Welfare, Tokyo; Chief, Department of Epidemiology, Institute of Public Health, Tokyo, *808 Den'en-Chofu 2-Chome, Ohta-ku, Tokyo, Japan.*

Strodtbeck, F. L., Ph.D. (Harvard Univ.), Assistant Professor of Sociology, Yale University, *576 Whalley Avenue, New Haven 11, Connecticut.*

Taranger, Aksel, B.Sc. (Univ. of Wis.), General Sales Manager, Norsk Aluminium Company, Loekkeveien 9, Oslo, Norway.

van Dantzig, D., Ph.D. (Groningen, Netherlands), Head of Department of Statistics, Mathematical Centre, Amsterdam; Professor, University of Amsterdam, *Valeriusstraat 58, Amsterdam-Zuid, Netherlands.*

Varnum, E. C., M.S. (Univ. of Mich.), Mathematician, Barber-Colman Company, Rockford, Illinois.

REPORT OF THE OAK RIDGE MEETING OF THE INSTITUTE

The forty-sixth meeting of the Institute of Mathematical Statistics was held jointly with the Biometric Society (Eastern North American Region) at the

Oak Terrace in Oak Ridge, Tennessee, on March 15–17, 1951. The registration was one hundred four, including the following members of the Institute:

G. E. Albert, Kenneth J. Arnold, Joseph Berkson, R. C. Bose, R. A. Bradley, A. E. Brandt, Irwin J. Bross, Lyle D. Calvin, Osmer Carpenter, Jerome Cornfield, Phelps P. Crump, Edward E. Cureton, D. B. Duncan, Arthur M. Dutton, Meyer Dwass, Churchill Eisenhart, Evelyn Fix, B. G. Greenberg, Samuel W. Greenhouse, Robert J. Haden, Boyd Harshbarger, Alston S. Householder, Oscar Kempthorne, Allyn W. Kimball, Jacob E. Lieberman, Joseph Mandelson, Nathan Mantel, Margaret P. Martin, Herbert A. Meyer, Jack Moshman, T. Ellison Neal, H. W. Norton, I. Olkin, D. A. Probst, John H. Reynolds, S. N. Roy, Marvin Schneiderman, Esther Seiden, H. Fairfield Smith, Melvin D. Springer, William F. Taylor, D. Teichroew, M. E. Terry, Donovan J. Thompson, David V. Tiedeman, John W. Tukey, Myron J. Willis, W. J. Youden.

Professor Paul Densen, University of Pittsburgh, presided at the first Institute session, Thursday afternoon, March 15, devoted to *Public Health Statistics* Papers read were:

1. *A Simple Stochastic Model of Recovery, Relapse, Death and Loss of Patients*. Evelyn Fix and Jerzy Neyman, University of California, Berkeley.
2. *An Elementary Stochastic Process for a Syphilis Population*. B. G. Greenberg, University of North Carolina.

The Oak Ridge National Laboratory acted as host at a smoker and beer party in honor of the visitors at the Ridge Recreation Hall on Thursday evening. About seventy-five people were present.

On Friday morning Dr. W. A. Arnold, Oak Ridge National Laboratory, acted as chairman at a session on *Bioassay with Quantal Responses*. Papers presented were:

1. *Why I Prefer Logits to Probits and Sinits*. Joseph Berkson, Mayo Clinic.
2. *How Much Does the Choice of Metameter Matter?* J. W. Tukey, Princeton University.
3. *Extensions of Elementary Methods in Bioassay*. Irwin Bross, The Johns Hopkins University.

Jerome Cornfield, National Institute of Health, initiated a lively discussion.

Professor C. L. Comar, University of Tennessee, was chairman of the meeting Friday afternoon concerning *Experimental Design*. The following papers were presented:

1. *Incomplete Block Designs*. R. C. Bose, University of North Carolina.
2. *Fractional Replication*. Oscar Kempthorne, Iowa State College.
3. *The Analysis of Long Term Experiments*. A. M. Dutton, Iowa State College.
4. *Testing-for-Preference Experiments*. L. D. Calvin, G. D. Searle and Company, Chicago.

About seventy people were present at a banquet sponsored by both organizations. Dr. E. R. McCrady, U. S. Atomic Energy Commission, acted as toastmaster and introduced Dr. A. M. Weinberg, Research Director of the Oak Ridge National Laboratory. Dr. Weinberg welcomed the visitors and spoke briefly of the analogy between nuclear physics and biometrics (in the broad sense) and stochastic processes in nuclear research.

The first session Saturday morning on *Multivariate Analysis* was presided over by Professor E. E. Cureton, University of Tennessee. Papers presented were:

1. *On the Properties and Statistical Purposes of Some Well-known and Some New Tests in Multivariate Analysis.* S. N. Roy, University of North Carolina.
2. *Some Applications of Compound Symmetry Tests.* A. W. Kimball, Oak Ridge National Laboratory.
3. *Some Preliminary Results of Multivariate Discriminant Analysis.* D. V. Tiedeman, Harvard University.

There followed two concurrent sessions of contributed papers. Dr. A. S. Householder, Oak Ridge National Laboratory, presided at the first session, at which the following papers were contributed:

1. *Confidence Intervals for the Mean Rate at Which Radioactive Particles Impinge on a Type I Counter. Preliminary Report.* G. E. Albert, University of Tennessee and Oak Ridge National Laboratory.
2. *A Problem of Elapsed Times in a Sequence of Events.* Osmer Carpenter, Carbide and Carbon Chemicals Division, Oak Ridge.
3. *On the Existence of Unbiased Tests for Testing Composite Hypotheses.* Esther Seiden, University of Buffalo.
4. *Group Divisible Incomplete Block Designs.* R. C. Rose, University of North Carolina.
5. *Orthogonal Arrays of Strength Two and Three.* R. C. Bose and Kenneth A. Bush, University of North Carolina.
6. *The Structure of Balanced Incomplete Block Designs, and the Impossibility of Certain Unsymmetrical Cases.* (By Title.) William S. Connor, University of North Carolina.
7. *Some Bounded Significance Level Tests for the Median.* (By Title.) John E. Walsh, The Rand Corporation.

Professor W. S. Snyder, University of Tennessee, presided at the second session Saturday morning, at which the following contributed papers were presented:

1. *Joint Sampling Distribution of the Mean and Standard Deviation for Distribution Functions of the First Kind.* Melvin D. Springer, U. S. Naval Ordnance, Indianapolis.
2. *On Certain Distribution Problems in Multivariate Analysis. Preliminary Report.* Ingram Olkin, University of North Carolina.
3. *A Unified Approach to a Wide Class of Distribution Problems in Multivariate Analysis.* S. N. Roy, University of North Carolina.
4. *An Extension of the Buffon Needle Problem.* Nathan Mantel, National Cancer Institute.
5. *A Generalization of Sampling without Replacement from a Finite Universe.* D. G. Horvitz and D. J. Thompson, Iowa State College.
6. *A Problem in Two-Stage Sampling.* (By Title.) B. M. Seelbinder, University of North Carolina.
7. *Bounds on a Distribution Which Are a Function of Moments to Order Four. Preliminary Report.* (By Title). Marvin Zelen, University of North Carolina.
8. *An Inconsistency among Type A Regions.* (By Title.) Herman Chernoff, University of Illinois.
9. *Stochastic Approximation. Preliminary Report.* (By Title.) Herbert Robbins and S. Monro, University of North Carolina.

Additional special events of the meeting included a tour through the American Museum of Atomic Energy and a tea for the ladies given by Mrs. A. W. Kimball.

JACK MOSHMAN
Assistant Secretary

PUBLICATIONS RECEIVED

BUTTERBAUGH, GRANT I., ***A Bibliography of Statistical Quality Control, Supplement***, **University of Washington Press, Seattle, 1951, 141 pp., $2.00.**

Tables d'Intérêts et d'Annuités, Crédit Communal de Belgique, Brussels, 1950, 163 pp.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION *March 1951* **Vol. 46 No. 253**

1108 16th Street, N. W., Washington 6, D. C.

THE INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

OFFICERS FOR 1951

President:

P. S. Dwyer, University of Michigan, Ann Arbor

President-Elect:

M. A. Girshick, Stanford University, Stanford

Secretary-Treasurer:

C. H. Fischer, University of Michigan, Ann Arbor

Editor:

T. W. Anderson, Columbia University, New York

The purpose of the Institute of Mathematical Statistics is to stimulate research in the mathematical theory of statistics and to promote cooperation between the field of pure research and the fields of application.

Membership dues including subscription to the Annals of Mathematical Statistics are $10.00 per year within the United States and Canada and $7.00 per year elsewhere. Dues and inquiries regarding membership in the Institute should be sent to the Secretary-Treasurer of the Institute.

MEETING OF THE INSTITUTE

SUMMER MEETING—MINNEAPOLIS, MINNESOTA—September 3-7, 1951.

To be held in conjunction with the American Mathematical Society and the Mathematical Association of America. Abstracts must be in the hands of Associate Secretary R. J. Arnold, North Hall, University of Wisconsin, Madison, Wisconsin, not later than July 15.